

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:15:08 ; Search time 68.88 Seconds
(without alignments)
64.941 Million cell updates/sec

Title: US-09-095-478A-8

Perfect score: 673

Sequence: 1 DFWGMMWNCNVIAMITRE.....VRKSHITGPLLIVHCTAGVGR 122

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268.5	39.9	2485	1 PTND_HUMAN	Q12923 homo sapien
2	266.5	39.6	377	1 PTP2_DICDI	P34138 dictyosteli
3	250	37.1	913	1 PTN3_HUMAN	P26045 homo sapien
4	244	36.3	2316	1 PTP2_RAT	Q62656 rattus norv
5	239	35.5	1337	1 PTPJ_HUMAN	Q12913 homo sapien
6	238	35.4	2314	1 PTP2_HUMAN	P23471 homo sapien
7	236.5	35.1	1452	1 PTM5_HUMAN	P28827 homo sapien
8	235.5	35.0	1152	1 CD45_MOUSE	P06800 mus musculu
9	235.5	35.0	1452	1 PTPM_MOUSE	P28828 mus musculu
10	233.5	34.7	1255	1 CD45_RAT	P04157 rattus norv
11	232	34.5	1187	1 PTNE_HUMAN	Q15678 homo sapien
12	232	34.5	1238	1 PTPJ_MOUSE	Q64455 mus musculu
13	232	34.5	1912	1 PTPD_HUMAN	P23468 homo sapien
14	231	34.3	1997	1 PTPD_HUMAN	P23467 homo sapien
15	230.5	34.2	1304	1 CD45_HUMAN	P08575 homo sapien
16	230.5	34.2	1430	1 PTPU_HUMAN	Q92729 homo sapien
17	230	34.2	1897	1 PTPF_HUMAN	P10586 homo sapien
18	229.5	34.1	382	1 PTPF_MOUSE	Q06180 mus musculu
19	229	34.0	2029	1 LAR_DROME	P16621 drosophila
20	228.5	34.0	1026	1 PTP1_CAEEL	P28191 caenorhabdi
21	225.5	33.5	363	1 PTN2_RAT	P35233 rattus norv
22	224.5	33.4	415	1 PTN2_HUMAN	P17706 homo sapien
23	224	33.3	775	1 PTNC_MOUSE	P35831 mus musculu
24	223.5	33.2	1176	1 PTNL_MOUSE	Q62136 mus musculu
25	222	33.0	1189	1 PTNE_MOUSE	Q62130 mus musculu
26	221.5	32.9	1705	1 PTPC_MOUSE	P70289 mus musculu
27	221	32.8	780	1 PTPC_HUMAN	Q05209 homo sapien
28	220.5	32.8	1439	1 PTPK_HUMAN	O15262 homo sapien
29	220.5	32.8	1457	1 PTPK_MOUSE	P29349 mus musculu
30	219	32.5	845	1 CSW_DROME	P29349 drosophila
31	218	32.4	434	1 PTN1_CHICK	O13016 gallus gall
32	218	32.4	435	1 PTNL_HUMAN	P18031 homo sapien
33	217.5	32.3	1174	1 PTNL_HUMAN	Q15825 homo sapien

34	217.5	32.3	1175	1 PTNL_RAT	Q62728 rattus norv
35	217.5	32.3	1445	1 PTPG_HUMAN	P23470 homo sapien
36	216	32.1	432	1 PTN1_RAT	P20417 rattus norv
37	215.5	32.0	1711	1 PTPC_RAT	Q64612 rattus norv
38	214.5	31.9	926	1 PTN4_HUMAN	P20074 homo sapien
39	214	31.8	432	1 PTN1_MOUSE	P35821 mus musculu
40	210.5	31.3	593	1 PTN9_HUMAN	P43378 homo sapien
41	207.5	30.8	979	1 PTPN_HUMAN	Q16849 homo sapien
42	207	30.8	360	1 PTN7_MOUSE	P29336 homo sapien
43	206	30.6	595	1 PTN6_MOUSE	P29351 mus musculu
44	205.5	30.5	979	1 PTPN_BOVIN	P56722 bos taurus
45	205.5	30.5	979	1 PTPN_MOUSE	Q60673 mus musculu

ALIGNMENTS

RESULT	1
PTND_HUMAN	STANDARD; PRT; 2485 AA.
AC	Q12923; Q15263; Q16826; Q15264; Q15265; Q15159;
DT	20-AUG-2001 (Rel. 40, Created)
DT	20-AUG-2001 (Rel. 40, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 13 (EC 3.1.3.48)
DE	(PROTEIN-TYROSINE PHOSPHATASE 1E) (PTP-EL) (PTP-BAS) (PROTEIN-TYROSINE PHOSPHATASE PTP1) (FAS-ASSOCIATED PROTEIN-TYROSINE PHOSPHATASE 1)
DE	(FAP-1).
GN	PTPN13 OR PTPLE OR PTP1 OR PNP1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Breast Carcinoma;
RX	MEDLINE=94350988; PubMed=8071359;
RA	Bajville D., Ahmad S., Stocco R., Shen S.-H.;
RT	"A novel protein-tyrosine phosphatase with homology to both the cytoskeletal proteins of the band 4.1 family and junction-associated quanylate kinases."
RL	J. Biol. Chem. 269:22320-22327(1994).
RN	[2]
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC	TISSUE=Leukemia;
RX	MEDLINE=94116679; PubMed=8287977;
RA	Maekawa K., Imagawa N., Nagamatsu M., Harada S.;
RT	"Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane-binding domain and GLGF repeats."
RL	FEBS Lett. 337:200-206(1994).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Fibroblast;
RX	MEDLINE=95014139; PubMed=7929060;
RA	Sayas J., Classon-Welsh L., Heldin C.-H., Genez L.J.;
RT	"Cloning and characterization of PTP1, a protein tyrosine phosphatase with similarities to cytoskeletal-associated proteins."
RL	J. Biol. Chem. 269:24082-24089(1994).
RN	[4]
RP	SEQUENCE OF 1216-2490 FROM N.A.
RC	TISSUE=Pancreas;
RA	Wang H.-Y.;
RL	Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RN	[5]
RP	STRUCTURE BY NMR OF 1361-1456.
RX	MEDLINE=20170882; PubMed=10704206;
RA	Kozlov G., Gehring K., Ekiel I.;
RT	"Solution structure of the PDZ domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor."
RL	Biochemistry 39:2572-2580(2000).
CC	FUNCTION: BINDS TO A NEGATIVE REGULATORY DOMAIN IN FAS THAT INHIBITS FAS-INDUCED APOPTOSIS.

CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN
 CC TYROSINE + PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES WITH THE EXCEPTION OF
 CC THE LIVER AND SKELETAL MUSCLE. MOST ABUNDANT IN LUNG, KIDNEY AND
 CC FETAL BRAIN.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U12128; AAB60339.1; -
 CC EMBL: D21209; BAA04750.1; -
 CC EMBL: D21210; BAA04751.1; -
 CC EMBL: D21211; BAA04752.1; -
 CC EMBL: X80289; CAA56563.1; -
 CC EMBL: X79676; CAA56124.1; -
 CC PDB: 3PDZ; 17-MAR-00.
 CC MIM: 600267; -
 CC InterPro: IPR000299; Band_4.1.
 CC InterPro: IPR001478; PDZ.
 CC InterPro: IPR000387; Tyr_phosphatase.
 CC InterPro: IPR000242; Tyr_prot_phptase.
 CC Pfam: PF00373; Band_41; 1.
 CC Pfam: PF00595; PDZ; 5.
 CC Pfam: PF00102; Y_phosphatase; 1.
 CC PRINTS: PR00700; PRTYRPHPTASE.
 CC PRINTS: PR00935; BAND41.
 CC SMART: SM00295; B41; 1.
 CC SMART: SM00228; PDZ; 5.
 CC SMART: SM00194; PTPC; 1.
 CC PROSITE: PS00660; BAND_41_1; FALSE_NEG.
 CC PROSITE: PS00661; BAND_41_2; FALSE_NEG.
 CC PROSITE: PS50057; BAND_41_3; 1.
 CC PROSITE: PS50106; PDZ; 5.
 CC PROSITE: PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
 CC PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 CC PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 CC Structural protein; Cytoskeleton; Hydrolase; Repeat; 3D-structure;
 CC Alternative splicing; Coiled coil.
 CC KW DOMAIN 56 59 POLY-LEU.
 CC FT DOMAIN 585 879 BAND 4.1-LIKE.
 CC FT DOMAIN 2237 2485 PROTEIN-TYROSINE PHOSPHATASE.
 CC FT DOMAIN 379 399 COILED COIL (POTENTIAL).
 CC FT DOMAIN 469 504 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1775 1804 COILED COIL (POTENTIAL).
 CC FT DOMAIN 2057 2085 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1093 1178 PDZ 1.
 CC FT DOMAIN 1368 1452 PDZ 2.
 CC FT DOMAIN 1501 1588 PDZ 3.
 CC FT DOMAIN 1788 1868 PDZ 4.
 CC FT DOMAIN 1882 1965 PDZ 5.
 CC FT DOMAIN 1742 1749 POLY-SER.
 CC ACT_SITE 2408 2078 BY SIMILARITY.
 CC FT VARSPPLIC 884 1074 MISSING (IN ISOFORM 2).
 CC FT VARSPPLIC 1056 1074 MISSING (IN ISOFORM 3).
 CC FT FT FT LD -> FH (IN REF. 3).
 CC FT FT FT KDHWSRGTLRHIS -> DLSRSCHVYLAHL (IN
 CC FT REF. 4).
 CC FT GL -> A (IN REF. 4).
 CC FT S -> P (IN REF. 4).
 CC FT KP -> RS (IN REF. 4).
 CC FT T -> TVLFDK (IN REF. 1).
 CC FT 1238 1239
 CC FT 1357 1357
 CC FT 1362 1363
 CC FT 1383 1383

FT CONFLICT 1538 1538 P -> A (IN REF. 3).
 FT CONFLICT 1649 1649 R -> K (IN REF. 4).
 FT CONFLICT 1698 1714 KSOEDTCTMFYVPQKI -> RVKKIPFVPCFTILRR
 FT (IN REF. 4).
 FT CONFLICT 1797 1797 G -> A (IN REF. 3).
 FT CONFLICT 1856 1857 AA -> G (IN REF. 4).
 FT CONFLICT 2069 2069 A -> S (IN REF. 4).
 FT CONFLICT 2206 2210 GLDDQ -> VARS (IN REF. 4).
 SQ SEQUENCE 2485 AA; .276903 MW; 8D1B31597C66962B CRC64;
 Query Match 39.9%; Score 268.5; DB 1; Length 2485;
 Best Local Similarity 39.8%; Pred. No. 4.6e-21;
 Matches 49; Conservative 21; Mismatches 52; Indels 1; Gaps 1;
 QY .1 DFGMMNNCNVAMITREIEGGVVKCCSVVPVSL-KEPLEFKHFHVLLENFOITQVVF 59
 DB 2292 DFWQMIHQEQSTVAMTQVEGEKIKQRYWPNILKTTMVSNRLRLALYRMQLKGFV 2351
 QY 60 IRIFQIVKSTGKSHSVKHLOFINKPDHGTSPASVDFFIKYVYRVKSHITGPLLVHCTAG 119
 DB 2352 VRAMTLEDIQTREVRHISHLNFTAWPDHPTSPQDDLLTFISYMRHTRSGPIITHCSAG 2411
 QY 120 VGR 122
 DB 2412 IGR 2414
 RESULT 2
 PTP2_DICDI STANDARD; PRT; 377 AA.
 ID PTP2_DICDI
 AC P34138;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE 2 (EC 3.1.3.48) (PROTEIN-TYROSINE-
 DE PHOSPHATE PHOSPHODIOLASE 2) (PTPA).
 GN PTPB OR PTP2.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=94043028; PubMed=8226777;
 RA Ramalingam R., Shaw D.R., Ennis H.;
 RT "Cloning and functional expression of a Dictyostellium discoideum
 RT protein tyrosine phosphatase."
 RL J. Biol. Chem. 268:22680-22685(1993).
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L15420; AAA33242.1; -
 CC HSSP: P18031; 2HNP.
 CC Dictydb; DD05054; ptpb.
 CC InterPro: IPR000387; Tyr_phosphatase.
 CC InterPro: IPR000242; Tyr_prot_phptase.
 CC Pfam: PF00102; Y_phosphatase; 1.
 CC PRINTS: PR00700; PRTYRPHPTASE.
 CC SMART: SM00194; PTPC; 1.
 CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 CC PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 CC PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.

```
KW Hydrolase. 78 86 POLY-ASP.
FT DOMAIN 95 103 POLY-ASN.
FT ACT_SITE 281 BY SIMILARITY.
SQ SEQUENCE 377 AA; 43488 MW; 5B08F6EB54829FD9 CRC64;

Query Match 39.6%; Score 266.5; DB 1; Length 377;
Best Local Similarity 40.7%; Pred. No. 9.5e-22;
Matches 50; Conservative 20; Mismatches 50; Indels 3; Gaps 2;

QY 1 DFWMWNNCNCVIAITREIEGGVTKCCYWPVSLKEPLEKHEHVLLE-NFOITQYFV 59
Db 167 DFWKHEQNSLIIVLTREENFKTKDKYWD--KDEERYGNFIVKEDNITIPDILI 224
QY 60 IRIFQIVKSTGKSHVSKHLQIKWPDHGTPTASVDFFIKYVRYVRKSHITGPLLHCTAG 119
Db 225 RREFTLENLKDNRKIYHFQYTWPDHGTPTVSTGFLKFSVDFVDEKRSGLVIVHCSAG 284
QY 120 VGR 122
Db 285 IGR 287

RESULT 3
PTN3_HUMAN
ID PTN3_HUMAN STANDARD; PRT; 913 AA.
AC P26045;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE HI) (PTP-HI).
GN PTPN3 OR PTPH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE-91296738; PubMed-1648725;
RA Yang O., Tonks N.K.;
RT Isolation of a cDNA clone encoding a human protein-tyrosine
RT phosphatase with homology to the cytoskeletal-associated proteins
RT band 4.1, ezrin, and talin."
RL Proc. Natl. Acad. Sci. U.S.A. 88:5949-5953(1991).
RN [2]
SEQUENCE OF 194-896 FROM N.A.
RP TISSUE-Colon;
RX MEDLINE-92327504; PubMed-1626183;
RA Arimura Y., Hinoda Y., Itoh F., Takekawa M., Tsujisaki M., Adachi M.,
RA Imai K., Yachi A.;
RT "cDNA cloning of new protein tyrosine phosphatases in the human
RT colon."
RL Tumour Biol. 13:180-186(1992).
RN [3]
SEQUENCE OF 899-913 FROM N.A.
RP MEDLINE-95179278; PubMed-7874267;
RA Ikuta S., Itoh F., Hinoda Y., Toyota M., Makiguchi Y., Imai K.,
RA Yachi A.;
RT "Expression of cytoskeletal-associated protein tyrosine phosphatase
RT PTPH1 mRNA in human hepatocellular carcinoma."
RL J. Gastroenterol. 29:727-732(1994).
CC -!- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
CC CYTOSKELETON.
CC -!- CATALYTIC ACTIVITY: PROTEIN-TYROSINE PHOSPHATASE + H(2)O =
CC PROTEIN-TYROSINE + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
-----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; M64572; AAA35647.1; -
DR EMBL; S39392; AAB22439.2; -
DR EMBL; S76309; AAB33583.1; -
DR PIR; A41109; A41109.
DR HSP; F18031; IPTT.
DR MIM; 176877; -
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00102; Y_phosphatase; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00507; BAND_41_3; 1.
DR PROSITE; PS0106; PDZ_1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00506; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00505; TYR_PHOSPHATASE_PTP; 1.
KW Structural protein; Cytoskeleton; Hydrolase.
FT DOMAIN 84 241
FT DOMAIN 510 582
FT DOMAIN 670 913
FT ACT_SITE 842 842
FT ACT_SITE 842 842 BY SIMILARITY.
SQ SEQUENCE 913 AA; 104029 MW; 29A539ACDE2F1515 CRC64;

Query Match 37.1%; Score 250; DB 1; Length 913;
Best Local Similarity 37.9%; Pred. No. 1.6e-19;
Matches 47; Conservative 23; Mismatches 48; Indels 6; Gaps 3;

QY 2 FWGMWNNCNCVIAITREIEGGVTKCCYWPVSLKEPLEKHEHVLLE-NFOITQYFV 59
Db 728 FWQVVDQKLSLIVMLTTLTERTKCHQWP---DPPDMHGGFHIQCSDECTIAYV 784
QY 60 IRIFQIVKSTGKSHVSKHLQIKWPDHGTPTASVDFFIKYVRYVRKSHI-TGPLLVHCTA 118
Db 785 SREMLVTNTQTGEHTVTHLQYVAMPDHGIPDDSSDFEVNVRSLRVDSEPLVHCSA 844
QY 119 GVGR 122
Db 845 GIGR 848

RESULT 4
PTPZ_RAT
ID PTPZ_RAT STANDARD; PRT; 2316 AA.
AC Q62656; Q62621;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE ZETA PRECURSOR (EC 3.1.3.48) (R-PTP-
DE ZETA) (PHOSPHACAN) (3F8 CHONDROITIN SULFATE PROTEOGLYCAN) (3H1 KERATAN
DE SULFATE PROTEOGLYCAN).
GN PTPRZ1 OR PTPRZ OR PTPZ.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID:10116;
```


AC 012913;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE ETA PRECURSOR (EC 3.1.3.48) (R-PTP-ETA)
DE (DENSITY ENHANCED PHOSPHATASE-1) (DEP-1) (CD148 ANTIGEN).
GN PTPRJ OR DEPL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=95024024; PubMed=7937872;
RA Oesman A., Yang Q., Tonks N.K.;
RT "Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase,
is enhanced with increasing cell density.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9680-9684(1994).
CC !- FUNCTION: MAY CONTRIBUTE TO THE MECHANISM OF CONTACT INHIBITION OF
CELL GROWTH.
CC !- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC !- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC !- PTM: N- AND O-GLYCOSYLATED.
CC !- SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.
CC !- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC !- DATABASE: NAME=PROW; NOTE=CD guide CD148 entry;
WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd148.htm".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U10886; AAB36687.1; --
DR HSSP: P18052; LYFO.
DR MIM: 600925; --
DR InterPro: IPR001777; FN_III.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; Tyr_prot_phptase.
DR Pfam: PF00041; fn3; 5
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00700; PTPRJPHPTASE.
DR SMART: SM00060; PTPC; 1.
DR SMART: SM00194; PTPC; 6.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00566; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00555; TYR_PHOSPHATASE_PTP; 1.
KW Signal; Glycoprotein; Transmembrane; Repeat; Hydrolase.
FT SIGNAL 1 35
FT CHAIN 36 1337
FT DOMAIN 36 975
FT TRANSMEM 976 996
FT DOMAIN 997 1337
FT DOMAIN 119 199
FT DOMAIN 366 446
FT DOMAIN 454 532
FT DOMAIN 540 615
FT DOMAIN 626 710
FT DOMAIN 1065 1337
FT ACT_SITE 1239 1239
FT CARBOHYD 72 72
FT CARBOHYD 82 82
FT CARBOHYD 93 93
FT CARBOHYD 104 104
FT CARBOHYD 142 142
FT CARBOHYD 172 172
FT CARBOHYD 192 192
FT CARBOHYD 231 231
FT CARBOHYD 258 258

FT CARBOHYD 278 278
FT CARBOHYD 342 342
FT CARBOHYD 351 351
FT CARBOHYD 376 376
FT CARBOHYD 391 391
FT CARBOHYD 396 396
FT CARBOHYD 413 413
FT CARBOHYD 431 431
FT CARBOHYD 501 501
FT CARBOHYD 525 525
FT CARBOHYD 536 536
FT CARBOHYD 582 582
FT CARBOHYD 603 603
FT CARBOHYD 618 618
FT CARBOHYD 628 628
FT CARBOHYD 637 637
FT CARBOHYD 666 666
FT CARBOHYD 669 669
FT CARBOHYD 761 761
FT CARBOHYD 772 772
FT CARBOHYD 784 784
FT CARBOHYD 790 790
FT CARBOHYD 824 824
FT CARBOHYD 910 910
FT CARBOHYD 937 937
SQ SEQUENCE 1337 AA; 145985 MW; 5686DE6D1F64236E CRC64;
Query-Match 35.5%; Score 239; DB 1; Length 1337;
Best Local Similarity 38.1%; Pred. No. 3 8e-18;
Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;
QY I DFWGMWNNCNVIAITREIEGGVIKCSYWPVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 1122 DFWRMVWKNVYALIMLTKEVGRTRKCEYWP--SKQAQYGDITVAMTSEIVLPEWTI 1179
QY 61 RIFQIVKKSSTGSHSVKHLQFIKWPDPHGPASVDFFIKY--VR-VVRKSHITGPIILVHC 116
DB 1180 RDTVANIOTSESHPLRQHFHTSWPDHGVPTDITLLINRYLVRYDMKSPSPSPILVHC 1239
QY 117 TAGVGR 122
DB 1240 SAGVGR 1245
RESULT: 6
PTPZ_HUMAN
ID PTPZ_HUMAN STANDARD; PRT: 2314 AA.
AC P23471;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
ET 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE ZETA PRECURSOR (EC 3.1.3.48) (R-PTP-
ZETA).
GN PTPZ1 OR PTPRZ OR PTPZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain;
RX MEDLINE=92366472; PubMed=1323835;
RA "A human transmembrane protein-tyrosine-phosphatase, PTP zeta, is
expressed in brain and has an N-terminal receptor domain homologous
to carbonic anhydrases.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7417-7421(1992).
RN [2]
RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).
RC TISSUE=Brain stem;
RX MEDLINE=93252948; PubMed=8387522;
RA Levy J.B., Canoll P.D., Silvennoinen O., Barnea G., Morse B.,

RA Honegger A.M., Huang J.-T., Cannizzaro L.A., Park S.-H., Druck T.,
 RA Huebner K., Sap J., Ehrlich M., Musacchio J.M., Schlessinger J.;
 RT "The cloning of a receptor-type protein tyrosine phosphatase
 RL expressed in the central nervous system.";
 RL J. Biol. Chem. 268:10573-10581(1993).
 RN [3]
 RP SEQUENCE OF 1479-2091 FROM N.A. (LONG FORM).
 RC TISSUE=Liver;
 RX MEDLINE=91006018; PubMed=2170109;
 RA Krueger N.X., Streuli M., Saito H.;
 RT "Structural diversity and evolution of human receptor-like protein
 FT tyrosine phosphatases";
 RL EMBO J. 9:3241-3252(1990).
 RN [4]
 RP SEQUENCE OF 1749-1990 AND 2047-2280 FROM N.A. (LONG FORM).
 RC TISSUE=Brain stem;
 RX MEDLINE=90384936; PubMed=2169617;
 RA Kaplan R., Morse B., Huebner K., Croce C., Hawk R., Ravera M.,
 RA Ricca G., Jaye M., Schlessinger J.;
 RT "Cloning of three human tyrosine phosphatases reveals a multigene
 RT family of receptor-linked protein-tyrosine-phosphatases expressed in
 RT brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).
 CC -i- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF SPECIFIC
 CC DEVELOPMENTAL PROCESSES IN THE CNS.
 CC -i- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -i- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -i- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -i- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE CENTRAL NERVOUS
 CC SYSTEM, WHERE IT IS LOCALIZED IN THE PURKINJE CELL LAYER OF THE
 CC CEREBELLUM, THE DENTATE GYRUS, AND THE SUBEPENDYMAL LAYER OF THE
 CC ANTERIOR HORN OF THE LATERAL VENTRICLE. DEVELOPMENTALLY REGULATED
 CC IN THE BRAIN.
 CC -i- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE EUKARYOTIC-
 CC TYPE CARBONIC ANHYDRASE FAMILY.
 CC -i- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC -i- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -i- CAUTION: CALLED RTTPASE BETA IN REF.2 AND 4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M93426; AAA60225.1; -;
 DR EMBL; X54135; CAA38070.1; -;
 DR PIR; S12054; S12054.
 DR PIR; A46151; A46151.
 DR HSP; P18052; LVFO.
 DR MIM; 176891; -;
 DR InterPro; IPR001148; Carb_anhydrase.
 DR InterPro; IPR001777; FN_III.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR InterPro; IPR000242; Tyr_prot_phphatase.
 DR Pfam; PF00194; carb_anhydrase; 1.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00102; Y_phosphatase; 2.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR ProDom; PD000865; Carb_anhydrase; 1.
 DR SMART; SM00060; FN3; 1.
 DR SMART; SM00194; PTPc; 2.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
 KW Signal; Glycoprotein; Transmembrane; Hydrolase; Alternative splicing.
 FT SIGNAL 1 24 BY SIMILARITY.
 FT CHAIN 25 2314 PROTEIN-TYROSINE PHOSPHATASE ZETA.
 FT DOMAIN 25 1635 EXTRACELLULAR (POTENTIAL).
 FT

FT TRANSMEM 1636 1661 POTENTIAL.
 FT DOMAIN 1662 2314 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 34 302 CARBONIC-ANHYDRASE LIKE.
 FT DOMAIN 312 406 FIBRONECTIN TYPE-III
 FT DOMAIN 1744 1997 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 1998 2314 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT BINDING 587 587 CHONDROITIN SULFATE (POTENTIAL).
 FT BINDING 637 637 CHONDROITIN SULFATE (POTENTIAL).
 FT BINDING 997 997 CHONDROITIN SULFATE (POTENTIAL).
 FT BINDING 1548 1548 CHONDROITIN SULFATE (POTENTIAL).
 FT BINDING 1550 1550 CHONDROITIN SULFATE (POTENTIAL).
 FT ACT_SITE 1932 1932 CHONDROITIN SULFATE (POTENTIAL).
 FT SITE -2222 2222 BY SIMILARITY.
 FT CARBOHYD 105 105 ANCESTRAL ACTIVE SITE.
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1122 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1456 1456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1561 1561 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1617 1617 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPIC 755 1614 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1722 1728 MISSING (IN REF. 2).
 SQ SEQUENCE 2314 AA; 254528 MW; 77DBDEF40F5FB42 CRC64;
 Query Match 35.4%; Score 238; DB 1; Length 2314;
 Best Local Similarity 37.0%; Pred. No. 9e-18;
 Matches 50; Conservative 24; Mismatches 43; Indels 18; Gaps 5;
 QY 1 DFWGMWENNCNVITAMITREIEGVKCCSWPVSLEKPELFKHFHVLLENFQITQYFVI 60
 DB 1839 DFRMIWEHNVIVMTNLVVERKKRCDQWPADGSE--EYGNFLVTQKSVQVLAITYT 1866
 QY 61 RIFQI-----VKKSTGKSHS-----VKHLQFIKWPDPGPPASVDFEIKYVRVVRKS-----H 107
 DB 1867 RNFTLRNTRKIKGSKGRPSGRVVTVQHYTCWPDMGVP---EYSLPVLTFVRKAAAYAKRH 1923
 QY 108 ITGPLLHVHCTAGVGR 122
 DB 1924 AVGPVVVHCSAGVGR 1938
 RESULT . 7
 PTPM_HUMAN STANDARD; PRT; 1452 AA.
 ID PTPM_HUMAN AC P28827;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE MU PRECURSOR (EC 3.1.3.48) (R-PTP-MU).
 GN PTPRM OR PTPRL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92008644; PubMed=1655529;
 RA Gebbink M.F.B.G., van Etten I., Hateboer G., Suijkerbuijk R.,
 RA Beijersbergen R., Geurts van Kessel A., Moolenaar W.H.;
 FT "Cloning, expression and chromosomal localization of a new putative


```
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.  
OR PROSITE; PS00740; MAM_1; 1.  
DR PROSITE; PS50060; MAM_2; 1.  
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;  
KW Immunoglobulin domain; Repeat.  
FT SIGNAL 1 20  
FT CHAIN 21 1452  
FT DOMAIN 21 1452  
FT TRANSMEM 21 742  
FT DOMAIN 743 764  
FT DOMAIN 765 1452  
FT DOMAIN 22 184  
FT DOMAIN 199 267  
FT DOMAIN 287 374  
FT DOMAIN 383 466  
FT DOMAIN 486 571  
FT DOMAIN 589 671  
FT DOMAIN 923 1153  
FT DOMAIN 1213 1447  
FT ACT_SITE 1095 1095  
FT ACT_SITE 1389 1389  
FT DISULFID 206 260  
FT CARBOHYD 72 92  
FT CARBOHYD 92 92  
FT CARBOHYD 131 131  
FT CARBOHYD 249 249  
FT CARBOHYD 406 406  
FT CARBOHYD 414 414  
FT CARBOHYD 454 454  
FT CARBOHYD 534 534  
FT CARBOHYD 544 544  
FT CARBOHYD 598 598  
FT CARBOHYD 651 651  
FT CARBOHYD 681 681  
SQ SEQUENCE 1452 AA; 163594 MW; F99D0FC67922CFC6; 1452;  
  
Query Match 35.0%; Score 235.5; DB 1; Length 1452;  
Best Local Similarity 36.3%; Pred. No. le-17;  
Matches 45; Conservative 20; Mismatches 54; Indels 5; Gaps 2;  
  
QY 1 DFWGMWNNCNVIAITREIEGGVKKCSYWPVSLKEPKEKHPHLLLENFOITQYFVI 60  
DB 981 DFWRMVHENTASITMVLNVEGRVKCKYWP---DTEIYKDIKVLIDTELLAEYVI 1037  
  
QY 61 RTIQVKKSTGKSHVKHLQFIKWDHGTGPASVDFFIYKVRVKSH--ITGPLLHVCTA 118  
DB 1038 RTFAVEKRGRIEIRQFHFETGPDHGVYPYHATGLLGFEVRQVSKSPNAGPLVVHCSA 1097  
  
QY 119 GVGR 122  
DB 1098 GAGR 1101  
  
RESULT 10  
CD45_RAT  
ID CD45_RAT STANDARD; PRT: 1255 AA.  
AC P04157;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE LEUKOCYTE COMMON ANTIGEN PRECURSOR (EC 3.1.3.48) (L-CA) (CD45) (T200)  
DE (FRAGMENT).  
GN PTPRC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI_TaxID=10116;  
[1]  
RP SEQUENCE FROM N.A.  
RA Barclay A.N., Jackson D.I., Willis A.C., Williams A.F.;  
RL Submitted (MAY-1987) to the EMBL/GenBank/DBJ databases.  
RN [2]
```

FT ACT_SITE 1119 1119 BY SIMILARITY
 FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 12 53 MISSING (IN ISOFORM 3).
 FT VARSPLIC 12 102 MISSING (IN ISOFORM 2).
 FT VARSPLIC 53 143 MISSING (IN ISOFORM 1).
 FT VARSPLIC 103 143 MISSING (IN ISOFORM 3).
 FT CONFLICT 38 38 S -> R (IN REF. 3).
 SQ SEQUENCE 1255 AA: 141208 MW: C257CBD2A355BCEA CRC64;

Query Match 34.7%; Score 233.5; DB 1; Length 1255;
 Best Local Similarity 37.6%; Pred. No. 1.4e-17;
 Matches 47; Conservative 17; Mismatches 58; Indels 3; Gaps 2;
 QY 1 DFWGMMWNNCNVIAITREIEGGVKKCSYWPVSLKEPFLKHFHVLLENFOITQYFVI 60
 Db 586 DFWKMIHQKATVIVVTRCEEGNKNCAEYPCMEGRTFRDVRVVVINDHKRCPYII 745
 QY 61 RIFQIV-KKSTGSHSVKHLQFIKWDGTPASVDFFIKYRVYR--KSHITGPLLHCT 117
 Db 746 QKLSIAHKKKATGREVTHIQFTSWPDHGVDPDHLKLLKRRVNAFNFSGPIVHCS 805
 QY 118 AGVGR 122
 Db 806 AGVGR 810

RESULT 11
 PTNE_HUMAN
 ID PTNE_HUMAN STANDARD; PRT; 1187 AA.
 AC Q15678;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 14 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE PEZ).
 GN PTPN14 OR PEZ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=95251727; PubMed=7733990;
 RA Smith A.L., Mitchell P.J., Shipley J., Gusterson B.A., Rogers M.V.,
 RA Crompton M.R.;
 RT "Pez: a novel human cDNA encoding protein tyrosine phosphatase- and
 RT ezrin-like domains.";
 RL Biochem. Biophys. Res. Commun. 209:959-965(1995).
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF HUMAN TISSUES
 CC INCLUDING KIDNEY, SKELETAL MUSCLE, LUNG AND PLACENTA.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use, by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X82676; CAA57993.1; .
 DR HSSP; P18052; LYFO.
 DR MIM; 603155;
 DR InterPro; IPR000299; Band_4.1.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR InterPro; IPR000242; Tyr_prot_phptase.
 DR Pfam; PF00373; Band_41; 1.
 DR Pfam; PF00102; Y-phosphatase; 1.
 DR PRINTS; PR00700; PRTYPHPHTASE.
 DR PRINTS; PR00935; BAND41.
 DR SMART; SM00295; B41; 1.
 DR SMART; SM00194; PTPC; 1.
 DR PROSITE; PS00660; BAND_41_1; 1.
 DR PROSITE; PS00661; BAND_41_2; 1.
 DR PROSITE; PS0057; BAND_41_3; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE; PS00556; TYR_PHOSPHATASE_2; 1.
 KW Structural protein; Cytoskeleton; Hydrolase.
 FT DOMAIN 75 239 BAND 4.1-LIKE.
 FT ACT_SITE 1121 1121 BY SIMILARITY.
 FT DOMAIN 566 573 POLY-PRO.
 FT DOMAIN 709 709 POLY-GLU.
 SQ SEQUENCE 1187 AA: 135239 MW: 015760B75E3574E3 CRC64;

Query Match 34.5%; Score 232; DB 1; Length 1187;
 Best Local Similarity 38.5%; Pred. No. 1.9e-17;
 Matches 55; Conservative 14; Mismatches 46; Indels 28; Gaps 5;
 QY 1 DFWGMMWNNCNVIAITREIEGGVKKCSYWPVSLKEPFLKHFHVLLENFOITQYFVI 60
 Db 932 DFWQWVEQGVNVIANVTAEEGGRTKSHRYW-----KLGSKHSSATYCKFKVTKF-- 1044
 QY 61 RIFQIVKKST-----GKSHSVKHLQFIKWDGTPASVDFFIKY-----VRVVK 105
 Db 1045 RTDSVCYATTGLVKVHLLSGQERTVWHLQYTDWPHGCPEDVQVGLSLEYEIQSVRRHTN 1104
 QY 106 SHITG-----PLLHCTAGVGR 122
 Db 1105 SMLEGTKNRHPPIVHCSAGVGR 1127

RESULT 12
 PTPN14_MOUSE
 ID PTPN14_MOUSE STANDARD; PRT; 1238 AA.
 AC Q61455;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE ETA PRECURSOR (EC 3.1.3.48) (R-PTP-ETA)
 DE (H2TP BETA-LIKE TYROSINE PHOSPHATASE).
 GN PTPRJ OR BTP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WR1-LPR/LPR;
 RX MEDLINE=96140699; PubMed=8549806;
 RA Kuramochi S., Matsuda S., Saitoh T., Ohsugi M.,
 RA Yamamoto T.;
 RT "Molecular cloning and characterization of Byp, a murine

receptor-type tyrosine phosphatase similar to human DEP-1.";
FEBS Lett. 378:7-14(1996).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EVERY TISSUE EXAMINED.
CC -!- SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D45212; BAA08146.1; .
CC HSP: P18052; LYFO.
CC MGD: MGI:104574; Ptpn1.
CC InterPro: IPR001777; FN_III.
CC InterPro: IPR000387; TYR_phosphatase.
CC InterPro: IPR000242; Tyr_prot_phptase.
CC Pfam: PF00041; fn3; 6.
CC Pfam: PF00102; Y_phosphatase; 1.
CC PRINTS: PR00700; PTPPHPTASE.
CC SMART: SM00060; FN3; 5.
CC SMART: SM00194; PTPC; 1.
CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE: PS00566; TYR_PHOSPHATASE_2; 1.
CC PROSITE: PS00555; TYR_PHOSPHATASE_PTP; 1.
CC Signal: Glycoprotein; Transmembrane; Repeat; Hydrolase.
KW SIGNAL 1 28
FT CHAIN 29 1238
FT DOMAIN 29 876
FT TRANSMEM 877 897
FT DOMAIN 898 1238
FT DOMAIN 40 112
FT DOMAIN 121 260
FT DOMAIN 268 348
FT DOMAIN 356 434
FT DOMAIN 442 518
FT DOMAIN 529 608
FT DOMAIN 966 1238
FT ACT_SITE 1140 1140
FT CARBOHYD 62 62
FT CARBOHYD 78 78
FT CARBOHYD 85 85
FT CARBOHYD 90 90
FT CARBOHYD 110 110
FT CARBOHYD 114 114
FT CARBOHYD 145 145
FT CARBOHYD 164 164
FT CARBOHYD 173 173
FT CARBOHYD 182 182
FT CARBOHYD 198 198
FT CARBOHYD 207 207
FT CARBOHYD 244 244
FT CARBOHYD 253 253
FT CARBOHYD 267 267
FT CARBOHYD 278 278
FT CARBOHYD 313 313
FT CARBOHYD 317 317
FT CARBOHYD 333 333
FT CARBOHYD 366 366
FT CARBOHYD 379 379
FT CARBOHYD 398 398
FT CARBOHYD 403 403
FT CARBOHYD 437 437
FT CARBOHYD 452 452
FT CARBOHYD 488 488
FT CARBOHYD 506 506
FT CARBOHYD 538 538

FT CARBOHYD 572 572
FT CARBOHYD 576 576
FT CARBOHYD 662 662
FT CARBOHYD 668 668
FT CARBOHYD 685 685
FT CARBOHYD 691 691
FT CARBOHYD 725 725
FT CARBOHYD 811 811
FT CARBOHYD 838 838
SQ SEQUENCE 1238 AA; 136782 MW; 939479EDC8016835 CRC64;
Query Match 34.5%; Score 232; DB 1; Length 1238;
Best Local Similarity 35.4%; Pred. No. 2e-17;
Matches 45; Conservative 23; Mismatches 51; Indels 8; Gaps 3;
QY ...1 DFWGMWNNCNVIAITREIEGGVIKCSYWPVSILKEPLEKHFHVLLENFQITQYFVI 60
Db 1023 DFWAVWEKNYATVMTLKVCQGRTKCEYWP--SKQADYGDITVAMTSEVLPDEWTI 1080
QY ...61 RIFQIVKSKTSHSVKHLQFIKWPDHGTPASVDFEIKYRVYVRKSHI-----TGPLLVH 115
Db 1081 RDEVVKNMNSHPLRQPHFTSWPDHGVPDITDILLNF-RYLVRDYMKGIPESPILVH 1139
QY 116 CTAGVGR 122
Db 1140 CSAGVGR 1146
RESULT +13
ID PTPD_HUMAN STANDARD; PRT; 1912 AA.
AC P23468;
BT 01-NOV-1991 (Ref. 20, Created)
DT 01-OCT-1996 (Ref. 34, Last sequence update)
DT 20-AUG-2001 (Ref. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE DELTA PRECURSOR (EC 3.1.3.48) (R-PTP-
DE DELTA).
GN PTPRD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
RX MEDLINE-95204468; PubMed-7896816;
RA Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT "Molecular characterization of the human transmembrane protein-
RT tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT alternative human transmembrane protein-tyrosine phosphatase delta
RT isoforms.";
RL J. Biol. Chem. 270:6722-6728(1995).
RN [2]
RP SEQUENCE OF 390-1912 FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-91006018; PubMed-2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases.";
RL EMBO J. 9:3241-3252(1990).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
CC TISSUES DUE TO ALTERNATIVE SPLICING.
CC -!- PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
CC FROM THE TRANSMEMBRANE SEGMENT.
CC -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -


```
FT DOMAIN 905 992 FIBRONECTIN TYPE-III 11.
FT DOMAIN 993 1082 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1083 1170 FIBRONECTIN TYPE-III 13.
FT DOMAIN 1171 1268 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1269 1352 FIBRONECTIN TYPE-III 15.
FT DOMAIN 1353 1442 FIBRONECTIN TYPE-III 16.
FT DOMAIN 1443 1530 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 1531 1618 BY SIMILARITY
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 598 598 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 829 829 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1096 1096 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1185 1185 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1470 1470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1474 1474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1518 1518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1997 AA; 224267 MW; 691E99BA7A1515DD CRC64;

Query Match 34.3%; Score 231; DB 1; Length 1997;
Best Local Similarity 35.4%; Pred. No. 4.4e-17;
Matches 45; Conservative 27; Mismatches 49; Indels 6; Gaps 3;

QY 1 DFWMGMWNNCNVIAITREIEGGVTKCSYWPVSLKEPFEKHFHVLLENFQITQYFVI 60
DB 1785 DFWMGMWNNCNVIAITREIEGGVTKCSYWPVSLKEPFEKHFHVLLENFQITQYFVI 1843

QY 61 RIFQIVKSTGKSHS-VKHLQFIKWPDHGTPTASVDFFIKYVR----YVRKSHITGPLLHV 115
DB 1844 REFKICGEQDLDAHLIRHFHVTWPDHGVPTTQSLIQFVTRVDYINRSPGAGPTVVH 1903

QY 116 CTAGVGR 122
DB 1904 CSAGVGR 1910

RESULT 15
CD45_HUMAN
ID CD45_HUMAN STANDARD: PRT; 1304 AA.
AC P08575;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LEUKOCYTE COMMON ANTIGEN PRECURSOR (EC 3.1.1.3.48) (L-CA) (CD45 ANTIGEN) (T200).
DE (T200).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RX MEDLINE=88061067; PubMed=2824653;
RA Streuli M., Hall L.R., Saga Y., Schlossman S.F., Saito H.;
```

```
RT "Differential usage of three exons generates at least five different
RT mRNAs encoding human leukocyte common antigens.";
RL J. Exp. Med. 166:1548-1566(1987).
RN [2]
RP FUNCTION.
RX MEDLINE=89017162; PubMed=2845400;
RA Charbonneau H., Tonks N.K., Walsh K.A., Fischer E.H.;
RT "The leukocyte common antigen (CD45): a putative receptor-linked
RT protein tyrosine phosphatase.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7182-7186(1988).
RN [3]
RP MUTAGENESIS.
RX MEDLINE=90316093; PubMed=1695146;
RA Streuli M., Krueger N.X., Thal T., Tang M., Saito H.;
RT "Distinct functional roles of the two intracellular phosphatase like
RT domains of the receptor-linked protein tyrosine phosphatases LCA and
RL LAR.";
RL EMBO J. 9:2399-2407(1990).
CC 1.1. FUNCTION: REQUIRED FOR T-CELLS ACTIVATION THROUGH THE ANTIGEN
CC RECEPTOR, THE FIRST PTPASE DOMAIN HAS ENZYMIC ACTIVITY, WHILE
CC THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE
CC FIRST ONE.
CC 1.2. CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC 1.3. ALTERNATIVE PRODUCTS: AT LEAST 8 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC 1.4. PTM: HEAVILY N- AND O-GLYCOSYLATED.
CC 1.5. SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC 1.6. SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC 1.7. DATABASE: NAME-PROW; NOTE-CD guide CD45 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd45.htm".
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DB EMIL: Y00638; CAA68669.1;
DB HSP; P18052; IYFO.
DB GLYCOSUITEDB; P08575;
DB MIM; 151460;
DB InterPro; IPR001777; FN_III.
DB InterPro; IPR000387; TVR_phosphatase.
DB InterPro; IPR000242; Tyr_prot_phptase.
DB Pfam; PF00041; fn3; 2.
DB Pfam; PF00102; Y_phosphatase; 2.
DB PRINTS; PR00700; PRTYPHPTASE.
DB SMART; SM00060; FN3; 2.
DB SMART; SM00194; PTPC; 2.
DB PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DB PROSITE; PS00386; TYR_PHOSPHATASE_2; 2.
DB PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Glycoprotein; Transmembrane; Phosphorylation; B-cell; T-cell; Repeat;
KW Alternative splicing; Hydrolase; Signal.
FT CHAIN 1 23 LEUKOCYTE COMMON ANTIGEN
FT DOMAIN 24 575 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 576 597 POTENTIAL.
FT DOMAIN 598 1304 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 387 479 FIBRONECTIN TYPE-III 1.
FT DOMAIN 480 571 FIBRONECTIN TYPE-III 2.
FT DOMAIN 670 919 PROTEIN-TYROSINE PHOSPHATASE 1.
FT DOMAIN 961 1235 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 1167 1167
FT ACT_SITE 78 78
FT CARBOHYD 90 90
FT CARBOHYD 95 95
FT CARBOHYD 184 184
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MUTAGEN 851 851 C->S: LOSS OF ACTIVITY.
SQ SEQUENCE 1304 AA; 147253 MW; 1F357BC5632618B2 CRC64;

Query Match 34.2%; Score 230.5; DB 1; Length 1304;
Best Local Similarity 38.4%; Pred. No. 3.1e-17;
Matches 48; Conservative 15; Mismatches 59; Indels 3; Gaps 2;

QY 1 DFWGMMWNNCNIAMITREIEGGVTKCCSYWPSVLSKEPLEFKHEHVHLLLENFOITQYFYI 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 DFWRMIEQKATVIMVTRCEEGNRNKAERYWPSMBEETRAFGDVVVKINQHKRCPDYII 792

QY 61 RIFQIV-KKSTGKSHSVKHLQFIKWPDPHTPASVDFFIKYVRYVR--KSHITGPLLHCT 117
   || || || || || || || || || || || || || || || || || || || || || ||
Db 793 QKLNIVNKKKATGREVTHIQFTSWPDHGVDPEDPHLLKLRNRVNAFSPFFSGPIVVHCS 852

QY 118 AGVGR 122
   |||||
Db 853 AGVGR 857
```

Search completed: March 30, 2002, 08:26:05
Job time: 657 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 07:17:38 ; Search time 69.17 Seconds
(without alignments)
134.354 Million cell updates/sec

Title: US-09-095-478A-8
Perfect score: 673
Sequence: 1 DFWGMWENNCHVIAITRE.....VRKSHITGPLLHVCTAGVGR 122
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	529	78.6	398	2 T08716	protein-tyrosine-p
2	274.5	40.8	2450	2 S71625	protein-tyrosine-p
3	268.5	39.9	2294	2 I67630	protein tyrosine p
4	268.5	39.9	2466	2 I67629	protein tyrosine p
5	266.5	39.6	377	1 A48711	protein-tyrosine-p
6	264.5	39.3	2490	1 A54971	protein-tyrosine-p
7	261	38.8	1156	2 T23308	hypothetical prote
8	250	37.1	913	1 A41109	protein-tyrosine-p
9	244	36.3	256	2 A40169	protein-tyrosine-p
10	240.5	35.7	597	2 B53978	protein-tyrosine-p
11	239	35.5	583	2 S17671	protein-tyrosine-p
12	239	35.5	1337	1 I38670	protein-tyrosine-p
13	238	35.4	2314	1 A46151	protein-tyrosine-p
14	237	35.2	1499	2 I50212	protein-tyrosine-p
15	236.5	35.1	1452	1 S17669	protein-tyrosine-p
16	235.5	35.0	1291	1 A28334	protein-tyrosine-p
17	235.5	35.0	1452	1 S17670	protein-tyrosine-p
18	233.5	34.7	694	2 A53978	protein-tyrosine-p
19	233.5	34.7	1273	1 TDRIT	protein-tyrosine-p
20	232.5	34.5	1436	2 JC5290	leukocyte common a
21	232	34.5	1187	1 JC4155	protein-tyrosine-p
22	232	34.5	1238	2 S68700	HPPP beta-like tyr
23	232	34.5	1912	2 A56178	protein-tyrosine-p
24	231	34.3	1501	2 I58148	protein-tyrosine-p
25	231	34.3	1907	2 S50893	protein-tyrosine-p
26	231	34.3	1997	1 S12050	protein-tyrosine-p
27	230.5	34.2	1304	1 A46546	leukocyte common a
28	230	34.2	582	2 A57068	protein-tyrosine-p
29	230	34.2	1290	2 A56493	leucocyte common a

30	230	34.2	1897	1 TDHULK	leukocyte antigen-
31	230	34.2	1898	2 S46216	leukocyte antigen-
32	229.5	34.1	382	1 A38191	protein-tyrosine-p
33	229	34.0	2029	1 TDFELK	protein-tyrosine-p
34	228.5	34.0	624	2 T19630	hypothetical prote
35	228.5	34.0	1026	2 T19631	hypothetical prote
36	228.5	34.0	1200	2 T43148	probable protein-t
37	228	33.9	2051	2 T30938	receptor tyrosine
38	227	33.7	1262	1 B48758	protein-tyrosine-p
39	227	33.7	1496	1 A48758	protein-tyrosine-p
40	227	33.7	1863	2 S46217	protein-tyrosine-p
41	226	33.6	382	1 S48748	protein-tyrosine-p
42	226	33.6	775	2 S55345	protein-tyrosine-p
43	226	33.6	1437	2 T31093	probable protein-t
44	225.5	33.5	363	1 S14294	protein-tyrosine-p
45	224.5	33.4	387	1 A60345	protein-tyrosine-p

ALIGNMENTS

RESULT 1
T08716
protein-tyrosine-phosphatase homolog DKF2p566K0524.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
C:Accession: T08716
R:Ansorge, W.; Winkler, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16472
A:Accession: T08716
A:Molecule type: mRNA
A:Residues: 1-398 <ANS>
A:Cross-references: EMBL:AL050040
A:Experimental source: fetal kidney; clone DKF2p566K0524
C:Genetics:
A:Note: DKF2p566K0524.1
C:Superfamily: protein-tyrosine-phosphatase homology
F:161-379/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match. 78.6%; Score 529; DB 2; Length 398;
Best Local Similarity 74.6%; Pred. No. 2.3e-48;
Matches 91; Conservative 16; Mismatches 15; Indels 0; Gaps 0;
Qy 1 DFWGMWENNCHVIAITREIEGGVKKCSYWPVSLKEPKEKHFHVLLENFQITQYFVI 60
Db 216 DFWQVLENNCHVIAITREIEGGVKKCSYWPVSLKEPKEKHFHVLLENFQITQYFVI 275
Qy 61 RFIQVKKSTGKSHVKKHLOFKWPDHGTTPASVDFIKYRVYRKSHITGPLLHVCTAGV 120
Db 276 RMFQVKKSTGKSHVKKHLOFKWPDHGTTPASVDFIKYRVYRKSHITGPLLHVCTAGV 335

Qy 121 GR 122
Db 336 GR 337

RESULT 2

S71625
protein-tyrosine-phosphatase (EC 3.1.3.48) RIP - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C:Accession: S71625; S67987
R:Chida, D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe, F.E.S. Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very
A:Reference number: S71625; MUID:95145716
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAAL12158.1; PID:g1232104

A:Experimental source: strain DBA/2; cell line MEL 745A

R:Wolf, B.B.; Brown, M.D.

FEB5 Lett. 376, 177-180, 1995

A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound signal transduction

A:Reference number: S67987; MUID:96105375

A:Accession: S67987

A:Molecule type: protein

A:Residues: 1098-1102 <MOL>

A:Experimental source: submaxillary glands

C:Genetics:

A:Gene: Ptpn13

A:Map position: 5

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;

F:566-860/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1089-1165/Domain: GLGF domain homology <GLG1>

F:1361-1437/Domain: GLGF domain homology <GLG2>

F:1495-1574/Domain: GLGF domain homology <GLG3>

F:1769-1840/Domain: GLGF domain homology <GLG4>

F:1863-1937/Domain: GLGF domain homology <GLG5>

F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>

F:2374/Active site: Cys (phosphocysteine intermediate) #status predicted

F:2380/Binding site: substrate phosphate (Arg) #status predicted

Query Match 40.8%; Score 274.5; DB 2; Length 2450;

Best Local Similarity 41.3%; Pred. No. 1.4e-20;

Matches 52; Conservative 23; Mismatches 44; Indels 7; Gaps 3;

QY 1 DFWGMWNNCNVIAITREIEGVVKKCSYWPVSL-KEPLEKHFHVLLENFQITQYFV 59

DB 2258 DFWQMIWEQKSTVIAMTQVEGEKIKQRYWPNILGKTTMWSNRLALVRMQQLKGFV 2317

QY 60 IRIFQIVKKSSTGSHSVKHLQFIKWPDPHTPASVD---FFIKVYRVYRKSHITGPLLVC 116

DB 2318 RVYMALEDIOTGEVRHSHLNFTAMPDHTPSQPDLLTFTISYMRHRRS---GPVITHC 2374

QY 117 TAGVGR 122

DB 2375 SAGIGR 2380

RESULT 3

I67630

protein tyrosine phosphatase (PTP-BAS, type 3) - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 16-Jul-1999

C:Accession: I67630

R:Maekawa, K.; Inagawa, N.; Nagamatsu, M.; Harada, S.

FEB5 Lett. 337, 200-206, 1994

A:Title: Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane-binding domain

A:Reference number: I53483; MUID:94116679

A:Accession: I67630

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

A:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;

F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match

Best Local Similarity 39.9%; Score 268.5; DB 2; Length 2294;

Matches 49; Conservative 21; Mismatches 52; Indels 1; Gaps 1;

QY 1 DFWGMWNNCNVIAITREIEGVVKKCSYWPVSL-KEPLEKHFHVLLENFQITQYFV 59

DB 2101 DFWQMIWEQKSTVIAMTQVEGEKIKQRYWPNILGKTTMWSNRLALVRMQQLKGFV 2160

QY 60 IRIFQIVKKSSTGSHSVKHLQFIKWPDPHTPASVDFFIKVYRVYRKSHITGPLLVC 119

DB 2375 SAGIGR 2380

QY 117 TAGVGR 122

DB 2375 SAGIGR 2380

DB 2161 VRAMTLEDIOTREVRHSHLNFTAMPDHTPSQPDLLTFTISYMRHRRSGPIITHCSAG 2220

QY 120 VGR 122

DB 2221 IGR 2223

RESULT 4

I67629

protein tyrosine phosphatase (PTP-BAS, type 2) - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 16-Jul-1999

C:Accession: I67629

R:Maekawa, K.; Inagawa, N.; Nagamatsu, M.; Harada, S.

FEB5 Lett. 337, 200-206, 1994

A:Title: Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane-binding domain

A:Reference number: I53483; MUID:94116679

A:Accession: I67629

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2466 <RES>

A:Cross-references: GB:D21210; NID:g452191; PIDN:BAA04751.1; PID:g452192

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology

F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1354-1430/Domain: GLGF domain homology <GLG2>

F:2216-2437/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match

Best Local Similarity 39.8%; Score 268.5; DB 2; Length 2466;

Matches 49; Conservative 21; Mismatches 52; Indels 1; Gaps 1;

QY 1 DFWGMWNNCNVIAITREIEGVVKKCSYWPVSL-KEPLEKHFHVLLENFQITQYFV 59

DB 2273 DFWQMIWEQKSTVIAMTQVEGEKIKQRYWPNILGKTTMWSNRLALVRMQQLKGFV 2332

QY 60 IRIFQIVKKSSTGSHSVKHLQFIKWPDPHTPASVDFFIKVYRVYRKSHITGPLLVC 119

DB 2333 VRAMTLEDIOTREVRHSHLNFTAMPDHTPSQPDLLTFTISYMRHRRSGPIITHCSAG 2392

QY 120 VGR 122

DB 2393 IGR 2395

RESULT 5

A48711

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 2 - slime mold (Dictyostelium)

C:Species: Dictyostelium discoideum

C:Date: 02-Jun-1995 #sequence_revision 08-Mar-1996 #text_change 11-Jun-1999

C:Accession: A48711; A53774

R:Ramalingam, R.; Shaw, D.R.; Ennis, H.L.

J. Biol. Chem. 268, 22680-22685, 1993

A:Title: Cloning and functional expression of a Dictyostelium discoideum protein tyrosine phosphatase

A:Reference number: A48711; MUID:94043028

A:Accession: A48711

A:Molecule type: mRNA

A:Residues: 1-377 <RAM>

A:Cross-references: GB:I15420; NID:g290036; PIDN:AAA33242.1; PID:g290037

R:Howard, P.K.; Gampier, M.; Hunter, T.; Firtel, R.A.

Mol. Cell. Biol. 14, 5154-5164, 1994

A:Title: Regulation by protein-tyrosine phosphatase PTP2 is distinct from that by PTP1

A:Reference number: A53774; MUID:94309635

A:Accession: A53774

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA; DNA

A:Residues: 1-377 <HOW>

A:Cross-references: GB:I15420; NID:g290036; PIDN:AAA33242.1; PID:g290037

A:Note: The translation of the nucleotide sequence is not complete in this paper

C:Superfamily: Dictyostelium protein-tyrosine-phosphatase, nonreceptor type 2; proteol

C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase

F:114-336/Domain: protein-tyrosine-phosphatase homology #status atypical <PTP>

[illegible]

```

Db      66 DFWRMIEHNVEVITMTNLVEKGRRKCDQWPTDGS--EYGSFLVNQNKNQVLAYTV 123
Qy      61 RIFOI-----VKSTGCKSHS-----VKHLQFIKWPDHGTGPASVDFFIKVRYVRKS-----H 107
       |   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     124 RNFTLRNTKKIKSGSKGRSSGLTYQHYTQWDPMGYP---EYSLPLAFVFKRTAQAKRH 180
Qy      108 ITGPLLVHCTAGVGR 122
       |   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     181 AVGPVVVHCAGVGR 195
       |   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 10
B53978
A:Title: protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type PTPx10 - African clawed
C:Species: Xenopus laevis (African clawed frog)
C:Date: 25-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C:Accession: B53978
R:Del Vecchio, R.L.; Tonks, N.K.
J. Biol. Chem. 269, 19639-19645, 1994
A:Title: Characterization of two structurally related xenopus laevis protein tyrosine
A:Reference number: A53978; MUID:94308257
A:Accession: B53978
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-597 <DEL>
A:Cross-references: GB:L33099; NID:g495671; PIDN:AAA21728.1; PID:g495672
A:Experimental source: ovary
A>Note: sequence extracted from NCBI backbone (NCBIN:l49759, NCBIPI:l49760)
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 9; cellular retinaldehyde
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphat
F:38-227/Domain: cellular retinaldehyde-binding protein homology <CRB>
F:328-564/Domain: protein-tyrosine-phosphatase homology <PTP>
F:516/Active site: Cys (phosphocysteine intermediate) #status predicted
F:522/Binding site: substrate phosphate (Arg) #status predicted

Query Match          35.7%; Score 240.5; DB 2; Length 597;
Best Local Similarity 35.0%; Pred. No. 1.2e-17;
Matches 48; Conservative 21; Mismatches 53; Indels 15; Gaps 2;

Qy      1 DFQGMWENCNVTAMTRIEGGVIKCWSWPVSLEPLEKFHFHVLLNFQTQYFVI 60
       | | | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     386 DFRMVMWEQRVLIVMTVRIEGRIRKCGVYLPLEGASEDTGHTFIIRNIHIDLFDQFKL 445
Qy      61 RIFQIVKSTGKSHSVKHLOFIKPWDHGTGPASVDFFTKVYVRK-----S 106
       | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     446 THEVYNKQTDERSVAHYQTMSPDFGVPKSASAMLDQRSQKHQAQAVQNLMGWGTG 505

Qy      107 HITG-PLLNVHCTAGVGR 122
       |   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     506 HPAGPPIVVHCAGVGR 522

RESULT 11
SL7671
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type beta - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 15-Mar-1996 #text_change 23-Jul-1999
C:Accession: SL7671; S40287
R:Gebjink, M.F.B.G.; van Etten, I.; Hateboer, G.; Suijkerbuijk, R.; Beijersbergen, R.
FEBS Lett. 290, 123-130, 1991
A:Title: Cloning, expression and chromosomal localization of a new putative receptor-
A:Reference number: SL7669; MUID:92008644
A:Accession: SL7671
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-583 <GEB>
A:Cross-references: EMBL:X58289
X:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
Submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphat
A:Reference number: S40280
A:Accession: S40287

```

A:Molecule type: mRNA
A:Residues: 377-483, 'n', 485-486 <HEN>
A:Cross-references: EMBL:223056; NID:9438149; PIDN:CAA80591.1; PID:9438150
C:Superfamily: protein-tyrosine-phosphatase, receptor type beta; fibronectin type III repeat
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane protein
F:311-536/Domain: protein-tyrosine-phosphatase homology <PTP>
F:488/Active site: Cys (phosphocysteine intermediate) #status predicted
F:494/Binding site: substrate phosphate (Arg) #status predicted

Query Match 35.5%; Score 239; DB 2; Length 583;
Best Local Similarity 36.2%; Pred. No. 1.7e-17;
Matches 46; Conservative 26; Mismatches 49; Indels 6; Gaps 3;

QY 1 DFWGMWENNVCNVIAMITREIEGGVKKCSYMPVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 369 DFWKMAEQNVHINIVMTQCEKGRVKCDHYPAD-QDPLYGDLTLQMVSESVLPETWI 427
QY 61 RIFQIVKKSSTGSHS-VKHLQFIKWDHCTPASVDFFIKYVR---YVRKSHITGPLLHV 115
DB 428 REFKICSEQLDAHLIRHFHYTVMPDHGVPEPTOSLIQFVTRVDYINRSPGAGPSVVH 487
QY 116 CTAGVGR 122
DB 488 CSAGVGR 494

RESULT 12
138670
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type J precursor - human
N:Alternate names: density enhanced phosphatase-1 (DEP-1); protein-tyrosine phosphatase
C:Species: Homo sapiens (man)
C:Date: 01-Mar-1996 #sequence_revision 08-Mar-1996 #text_change 22-Jun-1999
C:Accession: 138670; 152599
R:Ostman, A.; Yang, Q.; Tonks, N.K.
Proc. Natl. Acad. Sci. U.S.A. 91, 9680-9684, 1994
A:Title: Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase, is enhanced
A:Reference number: 138670; MUID:95024024
A:Accession: 138670
A:Molecule type: mRNA
A:Residues: 1-1337 <RES>
A:Cross-references: EMBL:U10886; NID:g558754; PID:g558755
A:Experimental source: HeLa cells
R:Honda, H.; Inazawa, J.; Nishida, J.; Yazaki, Y.; Hirai, H.
Blood 84, 4186-4194, 1994
A:Title: Molecular cloning, characterization, and chromosomal localization of a novel protein-tyrosine phosphatase, and enhanced expression of this protein with increasing cell density suggests a role in cell growth
A:Reference number: 152599; MUID:95086212
A:Accession: 152599
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-216, 'LTVGRRAA', 225-260, 'G', 262-285, 'GTEGGDASNTERRSRA', 302, 'S', 304, 'TAPVHDE'
A:Cross-references: GB:D37781; NID:g633072; PIDN:BAA07035.1; PID:g633073
C:Comment: Enhanced expression of this protein with increasing cell density suggests a role in cell growth
C:Genetics:
A:Gene: GDB:PTPRJ
A:Cross-references: GDB:385040; QMIM:600925
A:Map position: 19q13.4-19q13.4
C:Function:
A:Description: catalyzes hydrolysis of peptidyl-phosphotyrosine to peptidyl-tyrosine and phosphate
C:Superfamily: protein-tyrosine-phosphatase, receptor type J; fibronectin type III repeat
C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane protein
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-1337/Product: protein-tyrosine-phosphatase, receptor type J #status predicted <MAT>
F:118-197/Domain: fibronectin type III repeat homology <3FNA>
F:206-283/Domain: fibronectin type III repeat homology <3FNB>
F:284-356/Domain: fibronectin type III repeat homology #status atypical <3FNC>
F:365-445/Domain: fibronectin type III repeat homology <3FND>
F:453-530/Domain: fibronectin type III repeat homology <3FNE>
F:539-617/Domain: fibronectin type III repeat homology <3FNF>
F:720-804/Domain: fibronectin type III repeat homology <3FNG>
F:972-988/Domain: transmembrane #status predicted <TMN>
F:1065-1287/Domain: protein-tyrosine-phosphatase homology <PTP>
F:72,82,93,104,142,172,192,231,258,278,342,351,376,391,396,413,431,501,525,536,582,603,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000

F:1239/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1245/Binding site: substrate phosphate (Arg) #status predicted

Query Match 35.5%; Score 239; DB 1; Length 1337;
Best Local Similarity 38.1%; Pred. No. 4.2e-17;
Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;

QY 1 DFWGMWENNVCNVIAMITREIEGGVKKCSYMPVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 1122 DFWRMVWEKNVAVIIMLTCKVEQGRTKCEYWP--SKQAQDYGDITVAMTSEIVLPEWTI 1179
QY 61 RIFQIVKKSSTGSHS-VKHLQFIKWDHCTPASVDFFIKYVR---YVRKSHITGPLLHV 116
DB 1180 RFTTKVNIOTSESHPLRQFHFTSWPDHGVPTDITLINFRLVRYDMKQSPSPILVHC 1239
QY 117 TAGVGR 122
DB 1240 SAGVGR 1245

RESULT 13
A46151
protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type zeta precursor - human
N:Alternate names: protein-tyrosine-phosphatase beta; protein-tyrosine-phosphatase x1
C:Species: Homo sapiens (man)
C:Date: 08-Dec-1993 #sequence_revision 08-Feb-1996 #text_change 22-Jun-1999
C:Accession: A46151; A46700; B36065; S12054
R:Krueger, N.X.; Saito, H.
Proc. Natl. Acad. Sci. U.S.A. 89, 7417-7421, 1992
A:Title: A human transmembrane protein-tyrosine-phosphatase, PTP zeta, is expressed in the brain
A:Reference number: A46151; MUID:92366472
A:Accession: A46151
A:Molecule type: mRNA
A:Residues: 1-2314 <KRU>
A:Cross-references: GB:M93426; NID:q190743; PIDN:AAA60225.1; PID:q190744
A:Experimental source: fetal brain
A:Note: sequence extracted from NCBI backbone (NCBIN:110851, NCBI:110852)
A:Note: sequence inconsistent with the nucleotide translation
R:Levy, J.B.; Canoll, P.D.; Silvennoinen, O.; Barnea, G.; Morse, B.; Honegger, A.M.; nger, J.
J. Biol. Chem. 268, 10573-10581, 1993
A:Title: The cloning of a receptor-type protein tyrosine phosphatase expressed in the brain
A:Reference number: A46700; MUID:93252948
A:Accession: A46700
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1721, 1729-2314 <LEV>
A:Experimental source: brainstem
A:Note: sequence extracted from NCBI backbone (NCBI:131344)
R:Kaplan, R.; Morse, B.; Huebner, K.; Croce, C.; Howk, R.; Ravera, M.; Ricca, G.; Jay
Proc. Natl. Acad. Sci. U.S.A. 87, 7000-7004, 1990
A:Title: Cloning of three human tyrosine phosphatases reveals a multigene family of receptor-type protein tyrosine phosphatases
A:Reference number: A36065; MUID:90384936
A:Accession: B36065
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1749-1990; 2047-2280 <KAP>
A:Cross-references: GB:M34668; NID:q190738
R:Krueger, N.X.; Streuli, M.; Saito, H.
EMBO J. 9, 3241-3252, 1990
A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatases
A:Reference number: S12049; MUID:91006018
A:Accession: S12054
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1479-2091 <KR2>
A:Cross-references: GB:X54135; NID:g35795; PIDN:CAA38070.1; PID:g930104
A:Gene: GDB:PTPRZ1; PTPR2; PTP2; HPTP2; PTP18; RPTPB
A:Cross-references: GDB:127353; OMIM:176891
A:Map position: 7q31.3-7q31.3
C:Function:

Tue Apr 2 09:43:02 2002

Search completed: March 30, 2002, 08:16:25
Job time: 3527 sec

us-09-095-478a-8.rpt

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 07:10:48 ; Search time 77.75 Seconds
(without alignments)
116.231 Million cell updates/sec

Title: US-09-095-478A-8
Perfect score: 673
Sequence: 1 DFVGMWNNCNCVIAITRE.....VRKSHITGTLVHCAGVGR 122

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	A_Geneseq_1101.*
1:	/SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	673	100.0	122	20	AAW89252 Rat PTP10.
2	566	84.1	405	20	AAW89251 Mouse PTP05 isoform
3	566	84.1	426	20	AAW89249 Mouse PTP05. Mus
4	566	84.1	463	20	AAW89250 Mouse PTP05 isoform
5	268.5	39.9	2466	16	AAW71498 Human protein tyro
6	268.5	39.9	2466	19	AAW75999 Intracellular prot
7	268.5	39.9	2466	21	AAW90272 Human PTP1 phosph
8	268.5	39.9	2485	21	AAW19343 Amino acid sequenc
9	250	37.1	913	18	AAW12522 Protein tyrosine p
10	250	37.1	913	20	AAW25156 Human PTPHi protei
11	240	35.7	240	13	AAW20745 Human R-PTPase bet

12	240	35.7	242	22	AAW59384 Human protein tyro
13	240	35.7	579	21	AAW19773 Mouse vascular-end
14	239.5	35.6	253	22	AAW59374 Murine protein tyr
15	239	35.5	1337	16	AAW85203 huDEP-1. Homo sap
16	238	35.4	2308	15	AAW57902 Human RPTP-beta.
17	238	35.4	2308	22	AAW59586 Human receptor-tyr
18	237	35.2	1911	16	AAW71726 Human PTP-OB. Hom
19	237	35.2	1911	18	AAW27225 Human protein tyro
20	237	35.2	1911	20	AAW94027 Human protein tyro
21	237	35.2	1911	22	AAU01459 Human protein tyro
22	236.5	35.1	623	22	AAW25675 Human protein sequ
23	232.5	34.5	1499	22	AAW25768 Human protein sequ
24	231	34.3	1501	16	AAW72858 Rat receptor type-
25	231	34.3	1997	21	AAW19774 Human protein tyro
26	230.5	34.2	1430	19	AAW49907 Human pancreatic c
27	230	34.2	442	21	AAW56372 Human prostate can
28	230	34.2	607	21	AAW81783 Human protein tyro
29	230	34.2	607	21	AAW56098 LAR tyrosine phosph
30	230	34.2	647	22	AAW23746 Human ESR encoded
31	230	34.2	647	22	AAU14379 Human novel protei
32	230	34.2	1897	21	AAW19712 Human protein tyro
33	230	34.2	1897	21	AAW81785 Human protein tyro
34	230	34.2	1897	21	AAW56100 LAR tyrosine phosph
35	230	34.2	1907	22	AAU14143 Human novel protei
36	226	33.6	253	22	AAW59383 Human protein tyro
37	226	33.6	775	21	AAW67250 Mouse protein tyro
38	225	33.4	260	22	AAW59367 Human protein tyro
39	225	33.4	260	22	AAW59388 Human protein tyro
40	224.5	33.4	415	12	AAW14114 Non-receptor linke
41	224.5	33.4	1436	19	AAW41361 Receptor protein t
42	224	33.3	245	22	AAW59389 Human protein tyro
43	224	33.3	1149	22	AAW1048 Human polypeptide
44	221.5	32.9	249	22	AAW59371 Human protein tyro
45	221	32.8	780	21	AAW67252 Human protein tyro

ALIGNMENTS

RESULT 1
AAW89252
ID AAW89252 standard; Protein; 122 AA.
XX
AC AAW89252;
XX j0-MAR-1999 (first entry)
DT
XX Rat PTP10.
DE
XX
XX
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
OS Rattus sp.
XX
PN WO9849317-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08439.
XX
PR 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Ohrust S, Peles E, Plowman GD;

```

XX WPI; 1999-009434/01.
DR N-PSDB; AAV811747.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents rat PTP10. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 122 AA;
SQ
Query Match 100.0%; Score 673; DB 20; Length 122;
Best Local Similarity 100.0%; Pred. No. 9.1e-79; Mismatches 0; Gaps 0;
Matches 122; Conservative 0; Indels 0;
QY 1 DFGMMWENNVCNVIAMITREIEGGVVKCCSYMPVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 1 dfwgmwnncnvcnvlamitreieggvkkccsympvslkeplekfhfvllefnqityfvi 60
QY 61 RFIQIVKSTGKSHVYKHLQFKWPDHGTGPASVDFFIKYVYVRKSHITGPTLLVHCTAGV 120
DB 61 rfiqivkstkshvshvkhqfkwpdhtgtpasvdfvikyvyvrkshltgptllvhctagv 120
QY 121 GR 122
DB 121 gr 122
RESULT 2
AAW89251
ID AAW89251 standard; Protein; 405 AA.
XX
XX AAW89251;
AC AAW89251;
XX
XX 10-MAR-1999 (first entry)
DT Mouse PTP05 isoform #2.
DE
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
OS
XX WO9849317-A2.
PN
XX
XX 05-NOV-1998.
PD
XX
XX 27-APR-1998; 98WO-US08439.
PF
XX
XX 23-OCT-1997; 97US-0063595.
PR
XX 28-APR-1997; 97US-0044428.
PR
XX 20-MAY-1997; 97US-0047222.

```

```

PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallat B;
PI Markby D, Onrust S, Peles E, Plowman GD;
XX
XX WPI; 1999-009434/01.
DR N-PSDB; AAV811746.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 158-160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 405 AA;
SQ
Query Match 84.1%; Score 566; DB 20; Length 405;
Best Local Similarity 83.6%; Pred. No. 2.5e-64;
Matches 102; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
QY 1 DFGMMWENNVCNVIAMITREIEGGVVKCCSYMPVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 244 dfwgmwnncnvcnvlamitreieggvkkccsympvslkeplekfhfvllefnqityfvi 303
QY 61 RFIQIVKSTGKSHVYKHLQFKWPDHGTGPASVDFFIKYVYVRKSHITGPTLLVHCTAGV 120
DB 304 rfvqivkstksgkvkhqfkwpdhtgtpasvdfvikyvyvrkshltgptllvhctagv 363
QY 121 GR 122
DB 364 gr 365
RESULT 3
AAW89249;
ID AAW89249 standard; Protein; 426 AA.
XX
XX AAW89249;
AC AAW89249;
XX
XX 10-MAR-1999 (first entry)
DT Mouse PTP05.
DE
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
OS
XX WO9849317-A2.
PN

```

```
XX 05-NOV-1998.
PD 27-APR-1998; 98WO-US08439.
PF
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
PA
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;
PI
XX WPI; 1999-009434/01.
DR N-PSDB; AAV81744.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
PT
PS Claim 2; Page 155-157; 193pp; English.
XX
XX The present invention describes isolated,, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival.
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
SQ Sequence 426 AA;

Query Match 84.1%; Score 566; DB 20; Length 426;
Best Local Similarity 83.6%; Pred. No. 2.6e-64;
Matches 102; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 1 DFWGMWNNNCVNIAMITREIEGGVKKCSYWPVSLKEPFEKHFHVLLENFQITQYFVI 60
Db 244 dfwqmvlecnncvniamitreiecgvikcysywpisakeplefehsvfletfhvtqyftv 303

Qy 61 RFIQVKKSTGKSHSVKHLQFIKWPDHGTPASVDFFIKYVYVRKSHITGPLLHVCTAGV 120
Db 304 rvfqlvkkstgskvcvkhqlfctkwpdhtgtpasadffikyvrvyrkshitgpllvhcsagv 363

Qy 121 GR 122
Db 364 gr 365

RESULT 4
ID AAW89250 standard; Protein; 463 AA.
XX
AC AAW89250;
XX
DT 10-MAR-1999 (first entry)
XX
DE Mouse PTP05 isoform #1.
XX
```

```
KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW parkinson's disease; Huntington's disease.
XX
OS Mus sp.
XX
PN WO9849317-A2.
XX
PD 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
FA
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;
PI
XX WPI; 1999-009434/01.
DR N-PSDB; AAV81745.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
PT
PS Claim 2; Page 157-158; 193pp; English.
XX
XX This present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
SQ Sequence 463 AA;

Query Match 84.1%; Score 566; DB 20; Length 463;
Best Local Similarity 83.6%; Pred. No. 2.9e-64;
Matches 102; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 1 DFWGMWNNNCVNIAMITREIEGGVKKCSYWPVSLKEPFEKHFHVLLENFQITQYFVI 60
Db 281 dfwqmvlecnncvniamitreiecgvikcysywpisakeplefehsvfletfhvtqyftv 340

Qy 61 RFIQVKKSTGKSHSVKHLQFIKWPDHGTPASVDFFIKYVYVRKSHITGPLLHVCTAGV 120
Db 341 rvfqlvkkstgskvcvkhqlfctkwpdhtgtpasadffikyvrvyrkshitgpllvhcsagv 400

Qy 121 GR 122
Db 401 gr 402

RESULT 5
AAW71498
```


AC AAW12522;
 XX 22-MAY-1997 (first entry)
 DT Protein tyrosine phosphatase that localises to focal adhesion.
 DE
 XX Protein tyrosine phosphatase; PTPH1; focal adhesion;
 KW protein tyrosine kinase; malignancy; cancer; gene therapy;
 KW retrovirus; vector.
 XX
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Region 30..357
 FT /note= "N-terminal region shows homology with the
 FT N-terminal regions of the talin family"
 FT Region 357..663
 FT /note= "central region includes sequences with
 FT features of sites of phosphorylation by
 FT casein kinase II and p34cdc2"
 FT Modified-site 372
 FT /label= Phosphorylation
 FT /note= "putative p34cdc2 phosphorylation site"
 FT Modified-site 381
 FT /label= Phosphorylation
 FT /note= "putative p34cdc2 phosphorylation site"
 FT Modified-site 424..428
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 438..442
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 489..492
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 514..517
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 543..547
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Modified-site 607..610
 FT /label= Phosphorylation
 FT /note= "putative casein kinase II phosphorylation
 FT site"
 FT Region 664..931
 FT /note= "C-terminal region shows homology to "
 FT US5595911-A.
 FT 21-JAN-1997.
 FT 14-MAR-1990; 90US-0494036.
 FT 01-MAR-1991; 91US-0663579.
 FT 14-MAR-1990; 90US-0494036.
 FT 16-AUG-1993; 93US-0107420.
 FT (COLD-) COLD SPRING HARBOR LAB.
 FT Tonks NK;
 FT WPI; 1997-107583/10.
 FT N-PSDB; AAT58627.
 FT DNA encoding protein tyrosine phosphatase - for gene therapy of
 FT cancer
 XX

PS Disclosure; Fig 1A-B; 12pp; English.
 XX
 CC A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the
 CC dephosphorylation of proteins in which tyrosyl residues have been
 CC phosphorylated through the action of a protein tyrosine kinase
 CC (PTK). It localises to focal adhesions, a major site of action of
 CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)
 CC ctd. from HeLa cells. The PTPH1 cDNA can be incorporated into a
 CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian
 CC in sufficient quantities to overcome or counteract PTK activity.
 CC Phosphorylation of tyrosine residues at abnormal levels is
 CC prevented or reversed, resulting in the prevention or reversal of
 CC malignancy of cells.
 XX
 SQ Sequence 913 AA;
 Query Match 37.1%; Score 250; DB 18; Length 913;
 Best Local Similarity 37.9%; Pred. No. 3.3e-23;
 Matches 47; Conservative 23; Mismatches 48; Indels 6; Gaps 3;
 QY 2 FWCMMNENNVAMITREIEGGVVKCCSWPVSLKPELEFKH--PHVLENFQITQYFV 59
 DB 728 fwqvwqdklsilvmtlttergrtkchqwp---dpdvmmhggfhiqcqsedctiayv 784
 QY 60 IRIFQIVKSTGKSHSVKHLQFIKWPDPGPASVDFFIKYRYVRKSHI-TGPLLVHCTA 118
 DB 785 sremlvtntqgeehvtlqyvwpdhgipddssdflefvyrvsrldvsepvlvhcsa 844
 QY 1-9 GVGR 122
 DB .845 g1gr 848
 RESULT 10
 AAY25156
 ID AAY25156 standard; Protein; 913 AA.
 XX
 AC AAY25156;
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE Human PTPH1 protein.
 XX
 KW PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;
 KW localisation; treatment; overexpression; oncogenic; cell transformation;
 KW prevention; phosphotyrosine; disease; malignant.
 XX
 OS Homo sapiens.
 FH Key' Location/Qualifiers
 FT Region 30..357
 FT /note= "region of homology to the N-terminal domain
 FT of band 4.1, ezrin and talin. This region is
 FT known to be important for localisation to focal
 FT adhesions"
 FT Modified-site 372
 FT /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site 381
 FT /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site 424..428
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 434
 FT /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site 438..442
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 489..492
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 514..518
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 543..547
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 607..610

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:16:17 ; Search time 1554.02 Seconds
(without alignments)
2212.744 Million cell updates/sec

Title: US-09-095-478A-4

Perfect score: 320

Sequence: 1 GAAATTAATGTAATGTTAT.....ATTACAGACCCCTCTTGT 320

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_othr:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	50	15.6	1842	12	AK017003	AK017003 Mus muscu
2	27	8.4	934	13	AQ744125	AQ744125 HS_5507_A
3	26	8.1	544	10	AF247150	AF247150 AF247150
4	20	6.2	406	10	AA633769	AA633769 ac27d11.s
5	20	6.2	550	10	AA129435	AA129435 zn85f02.s
6	20	6.2	849	13	AZ681898	AZ681898 ENTLJ27TR
7	20	6.2	850	13	AZ682025	AZ682025 ENTLR91TF
8	20	6.2	893	13	AZ689826	AZ689826 ENTLJ27TR
9	19	5.9	259	10	BE711384	BE711384 PM2-H7068
10	19	5.9	615	10	AW559472	AW559472 EST314520
11	19	5.9	666	13	AZ083136	AZ083136 RPI-23-3
12	18	5.6	136	10	AA230263	AA230263 ncl3a04.f

C 13	18	5.6	155	11	BF662254	BF662254 maa80e02.
C 14	18	5.6	164	10	AI784154	AI784154 th88d03.x
C 15	18	5.6	177	10	AV298638	AV298638 AV298638
C 16	18	5.6	190	11	BB439182	BB439182 BB439182
C 17	18	5.6	192	11	D45598	D45598 HUMGS02790
C 18	18	5.6	199	10	AW072956	AW072956 xa43c01.x
C 19	18	5.6	219	10	AI864084	AI864084 wg63c11.x
C 20	18	5.6	228	10	BB214089	BB214089 BB214089
C 21	18	5.6	235	10	AV320214	AV320214 AV320214
C 22	18	5.6	243	10	AI991947	AI991947 ws42f11.x
C 23	18	5.6	245	10	AV230933	AV230933 AV230933
C 24	18	5.6	252	10	AV326992	AV326992 AV326992
C 25	18	5.6	253	11	BC670484	BC670484 DRNBEG04
C 26	18	5.6	259	10	AA558169	AA558169 n127c11.s
C 27	18	5.6	272	10	AA884408	AA884408 am16a04.s
C 28	18	5.6	275	10	AA047439	AA047439 zk73b02.s
C 29	18	5.6	276	10	BB052509	BB052509 BB052509
C 30	18	5.6	279	10	BB045084	BB045084 BB045084
C 31	18	5.6	279	10	BB303287	BB303287 BB303287
C 32	18	5.6	279	10	BB490959	BB490959 BB490959
C 33	18	5.6	279	10	BB565548	BB565548 BB565548
C 34	18	5.6	281	10	BB386000	BB386000 BB386000
C 35	18	5.6	281	11	R07867	R07867 vf16d05.sl
C 36	18	5.6	284	10	BB345232	BB345232 BB345232
C 37	18	5.6	284	10	BB435655	BB435655 BB435655
C 38	18	5.6	285	10	BB092661	BB092661 BB092661
C 39	18	5.6	286	10	BB379014	BB379014 BB379014
C 40	18	5.6	289	10	BB316043	BB316043 BB316043
C 41	18	5.6	291	10	BB303480	BB303480 BB303480
C 42	18	5.6	296	11	F37150	F37150 HSPD35253.H
C 43	18	5.6	298	11	BF466708	BF466708 UI-M-CG0p
C 44	18	5.6	299	10	AA508465	AA508465 nh66c02.s
C 45	18	5.6	300	10	AV127357	AV127357 AV127357

ALIGNMENTS

RESULT 1

AK017003
LOCUS 1842 bp mRNA 05-JUL-2001
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933431F18, full insert sequence.
ACCESSION AK017003
VERSION AK017003.1 GI:12856041
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:lib.RIKEN full-length enriched mouse cDNA library, clone:4933431F18.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1842)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349536

REFERENCE 2 (bases 1 to 1842)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3 (bases 1 to 1842)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,


```
ORGANISM      Oryctolagus cuniculus
REFERENCE      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        1 (bases 1 to 544)
                Nikolaev,A.I., Dubovaya,V.I., Litvinov,D.Y., Poltarauk,A.B.,
                Panasolov,V.S., Ivanov,D.S., Amchenkova,A.M., Narovlyansky,A.N.,
                Panasyuk,A.F. and Turpaev,K.T.
TITLE          Isolation of nitric oxide-induced genes from chondrocyte cDNA
JOURNAL        Library using subtractive hybridization
COMMENT        Unpublished (2000)
                Contact: Turpaev K
                Engelhardt Institute of Molecular Biology
                Vavilova 32, Moscow, 117984, Russia
                Email: turpaev@genome.eimb.relarn.ru.
FEATURES      source
                1..544
                /organism="Oryctolagus cuniculus"
                /db_xref="taxon:9986"
                /clone="61"
                /clone_lib="Oryctolagus cuniculus femoral head 2 months
                old"
                /tissue_type="femoral head"
                /cell_type="primary articular chondrocyte"
                /dev_stage="2 months old"
                /note="breed: New Zealand White; obtained via subtractive
                hybridization of chondrocytes preincubated with 0.5 mM
                nitroglutathione vs. untreated cells"
BASE COUNT    173 a 102 c 103 g 166 t
ORIGIN
Query Match   8.1%; Score 26; DB 10; Length 544;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 ATGATAACCGAGAGATAGAGGTGG 50
      |||||||||||||||||||||||||
Db 335 ATGATAACCGAGAGATAGAGGTGG 310

RESULT 4
LOCUS      AA633769          406 bp      mRNA      EST      06-MAR-1998
DEFINITION ac27d11.s1 Stratiagene ovary (#937217) Homo sapiens cDNA clone
IMAGE:857685 3' similar to contains Alu repetitive element;contains
element P7R5 repetitive element ;, mRNA sequence.
ACCESSION   AA633769
VERSION     AA633769.1 GI:2556983
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE    Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      1 (bases 1 to 406)
              Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
              Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
              J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
              White,Y., Wylie,T., Waterston,R. and Wilson,R.
              WashU-NCI human EST Project
              Unpublished (1997)
              Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Possible reversed clone: polyt not found
              Insert Length: 659 Std Error: 0.00
              Seq primer: -40ml3 fwd. ET from Amersham
              High quality sequence stop: 406.
              Location/Qualifiers
FEATURES      source
                1..550
                /organism="Homo sapiens"
                /db_xref="GDB:4594051"
                /db_xref="taxon:9606"
                /clone="IMAGE:564987"
                /clone_lib="Stratiagene lung carcinoma 937218"
                /tissue_type="lung carcinoma"
                /cell_line="NCI-H69"
                /dev_stage="cell line NCI-H69"
                /lab_host="SOLR (kanamycin resistant)"
                /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
                dt. Small cell carcinoma cell line NCI-H69. Average
```

```
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:857685"
/clone_lib="Stratiagene ovary (#937217)"
/sex="female"
/dev_stage="49 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
Total ovary tissue, normal, caucasian. Average insert
size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTT 3'
BASE COUNT    93 a 105 c 120 g 88 t
ORIGIN
Query Match   6.2%; Score 20; DB 10; Length 406;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 AACACGAGAGATAGAGGTGG 49
      |||||||||||||||||||
Db 328 AACACGAGAGATAGAGGTGG 347

RESULT 5
LOCUS      AA129435          550 bp      mRNA      EST      27-NOV-1996
DEFINITION zn85f02.s1 Stratiagene lung carcinoma 937218 Homo sapiens cDNA clone
IMAGE:564987 3', mRNA sequence.
ACCESSION   AA129435
VERSION     AA129435.1 GI:1689201
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE    Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      1 (bases 1 to 550)
              Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
              Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
              M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
              B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
              Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,
              Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
              Generation and analysis of 280,000 human expressed sequence tags
              Genome Res. 6 (9), 807-828 (1996)
              9704478
              Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: -40ML3 fwd. from Amersham
              High quality sequence stop: 353.
              Location/Qualifiers
FEATURES      source
                1..550
                /organism="Homo sapiens"
                /db_xref="GDB:4594051"
                /db_xref="taxon:9606"
                /clone="IMAGE:564987"
                /clone_lib="Stratiagene lung carcinoma 937218"
                /tissue_type="lung carcinoma"
                /cell_line="NCI-H69"
                /dev_stage="cell line NCI-H69"
                /lab_host="SOLR (kanamycin resistant)"
                /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
                dt. Small cell carcinoma cell line NCI-H69. Average
```

insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor sequence: 5'-GAATTCGGCAGCAG 3', -3' adaptor sequence: 5'-CTCGAGTTTTTTTTTTTTTTT 3'-

BASE COUNT	155 a	113 c	127 g	151 t	4 others
ORIGIN					

```

Query Match          * 6.28; Score 20; DB 10; Length 550;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 AGGAAGACCCACATTACAGG 308
    |||||
Db 302 AGGAAGACCCACATTACAGG 321

```

6	14-DEC-2000
RESULT	GSS
AZ681898	
LOCUS	849 bp DNA
DEFINITION	ENTLJ27FR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION	AZ681898
VERSION	AZ681898.1 GI:11818939
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica.
ORGANISM	Entamoeba histolytica
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.
AUTHORS	1 (bases 1 to 849)
TITLE	Loftus, B., Van Aken, S. and Fraser, C.
JOURNAL	Determination of clone end sequences from Entamoeba histolytica
COMMENT	HMI:IMSS sheared DNA library
	Unpublished (2000)
	Contact: Brendan J Loftus

FEATURES

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Antib: encastriy:duj

Seq primer: M13-Reverse

Class: Shotgun

High quality sequence start: 15

High quality sequence stop: 821.

Location/Qualifiers

```

FEATURES
source
1. 849
Location/Qualifiers
  /organism="Entamoeba histolytica"
  /strain="HM1:IMSS"
  /db_xref="taxon:5759"
  /clone_lib="Entamoeba histolytica Sheared DNA"
  /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
  Institute for Genomic Research (TIGR), Rockville, MD.
  Genomic DNA isolated from broth cultures of E. histolytica
  using a method described by Clark and Diamond (Clark,
  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
  method for isolate identification. Exp. Parasitol.
  77:450.). The DNA was mechanically sheared to give a
  tight size distribution (~2 kb). The v + i method used for
  the library construction is described in detail in Smith,
  H.O. and Venter, J.C. (Making small insert libraries for
  whole genome shotgun sequencing projects. In Genome
  Sequencing: A Practical Approach, eds. M. Vaudin and B.
  Barrell, Oxford University Press, 1999)."
```

Query Match 6.2%; Score 20; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 469 AAATAATTGTAATGTTATTG 488

RESULT 7
AZ682025/C

LOCUS	AZ682025	850 bp	DNA	GSS	14-DEC-2000
DEFINITION	ENTLR91TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic DNA sequence.				

ACCESSION AZ682025
 VERSION AZ682025.1 GI:11819171
 KEYWORDS GSS.

RECORD	SOURCE	ORGANISM
955	Entamoeba histolytica.	Entamoeba
	Entamoeba histolytica	Eukaryota: Entamoebidae: Entamoeba

REFERENCE
AUTHORS
TITLE

JOURNAL
Unpublished (2000)
HM1:IMSS sheared DNA library
determination of clone end sequences from Enclomeda histology
library

COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

fax: 501 550 0260
Email: enta@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library
Seq primer: M13-Forward
Class: shotgun

High quality sequence start: 28
High quality sequence stop: 693.

FEATURES
SOURCE

```
1. 050
source
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
```

```
/db_xref="taxon:5759"
/clone_lib="Entamoeba
```

/note=Vector: pHS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

BASE COUNT	311 a	. 95 c	113 g	331 t
ORIGIN	Barell, Oxford University			

Query Match	6.2%	Score 20;	DB 13;	Length 850;
Best Local Similarity	100.0%;	Pred. No. 7.1;		
Matches 20;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY . 3 AAATAATTGTAATGTTATTG 22

Db 305 AAATAATTGTAATGTTATTG 286

RESULT 8
AZ689826/C

LOCUS	AZ689826	893 bp	DNA	GSS	14-DEC-2000
DEFINITION	ENTJ742TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.				

accession
 version
 keywords

SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE Eukaryota; Entamoebidae; Entamoeba.
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: enta@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 8
High quality sequence stop: 529.
Location/Qualifiers
1. .893
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/note="Vector: pHOSt1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."
BASE COUNT 296 a 101 c 134 g 362 t
ORIGIN

Query Match 6.2%; Score 20; DB 13; Length 893;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 AATAATGTTATGTTATG 22
DB 562 AATAATGTTATGTTATG 543
|||||

RESULT 9
BE711384/c
LOCUS BE711384 259 bp mRNA EST 12-SEP-2000
DEFINITION PM2-HT0686-250500-002-d09 HT0686 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE711384
VERSION BE711384.1 GI:10099649
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 259)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?Cl=st2-PM2-HT0686-250500-002-d09&t3=2000-05-25&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 259.
Location/Qualifiers
1. .259
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0686"
/dev_stage="Adult"
/note="Organ: head_neck; vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 75 a 33 c 55 g 96 t
ORIGIN

Query Match 5.9%; Score 19; DB 10; Length 259;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 CACATTACAGGACCCCTCC 316
DB 109 CACATTACAGGACCCCTCC 91
|||||

RESULT 10
AW559472/c
LOCUS AW559472 615 bp mRNA EST 07-SEP-2000
DEFINITION EST314520 DSIR Medicago truncatula cDNA clone pDSIR-1905, mRNA sequence.
ACCESSION AW559472
VERSION AW559472.1 GI:7204898
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 615)
AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.
TITLE ESTs from roots of Medicago truncatula after inoculation with Phytophthora medicaginis
JOURNAL Unpublished (1999)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name: W250758e; TIGR sequence name: MTBA187K; More information, including clone ordering, is available at.

'http://chrysie.tamu.edu/medicago'
Seq primer: Sknod (CTA gAA CTA gty GAT CC).

FEATURES

```

source
1. .615
  /organism="Medicago truncatula"
  /cultivar="genotype A17"
  /db_xref="taxon:3880"
  /clone_lib="pDSIR-1905"
  /clone_lib="DSIR"
  /tissue_type="roots infected with Phytophthora medicaginis"
  /dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"
  /lab_host="E. coli strain XLOLR"
  /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells. Note: EST may be of fungal origin."
BASE COUNT      187 a 122 c 126 g 180 t
ORIGIN
Query Match      5.9%; Score 19; DB 10; Length 615;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 CCAGACCATGGCACTCTG 250
      |||||
Db 279 CCAGACCATGGCACTCTG 261

RESULT 11
AZ083136/c
LOCUS      666 bp      DNA      GSS      08-MAY-2000
DEFINITION      Mus musculus genomic clone RPCI-23-34P9,
DNA sequence.
ACCESSION      AZ083136
VERSION      1
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 666)
AUTHORS      Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL      Unpublished (1999)
COMMENT      Other-GSSs: RPCI-23-34P9.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 34 row: P column: 9
Seq primer: T7
Class: BAC ends.
Location/Qualifiers

```

source

```

1. .666
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-34P9"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      194 a 117 c 88 g 267 t
ORIGIN

```

Query Match

Best Local Similarity 5.9%; Score 19; DB 13; Length 666;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 GATAACCAGAGATAGAA 45

Db 646 GATAACCAGAGATAGAA 628

RESULT 12

```

AA230263/c
LOCUS      136 bp      mRNA      EST      19-AUG-1997
DEFINITION      nc13a04.r1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007982, mRNA
sequence.
ACCESSION      AA230263
VERSION      1
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 136)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 740 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 91.
Location/Qualifiers

```

FEATURES

source

```

1. .136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1007982"
/clone_lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10 ,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with

```

FEATURES

Location/Qualifiers

an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krieman."

BASE COUNT 57 a 15 c 17 g 47 t
ORIGIN

Query Match 5.6%; Score 18; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAATAATTGTAATGTT 18
|||||

Db 80 GAAATAATTGTAATGTT 63

RESULT 13
BF662254/c 155 bp mRNA EST 20-DEC-2000
LOCUS
DEFINITION maa80e02.y1 Soares mouse 3NbMS Mus musculus cDNA clone
IMAGE:3823179 5', mRNA sequence.

ACCESSION BF662254
VERSION BF662254.1 GI:11927388
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 155)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1459531
Seq primer: -40RP from Gibco.

FEATURES
source
1..155
Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3823179"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCTGAAGTGGAGGCGCGCTGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 47 a 24 c 29 g 55 t
ORIGIN

Query Match 5.6%; Score 18; DB 11; Length 155;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 TTCAAAATGTGAAGAAGT 184
|||||

Db 76 TTCAAAATGTGAAGAAGT 59

RESULT 14
AI784154/c

LOCUS
DEFINITION t88d03.xl Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone

AI784154 164 bp mRNA EST 13-DEC-1999
IMAGE:2125733 3', mRNA sequence.

ACCESSION AI784154
VERSION AI784154.1 GI:5325963
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 418 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 159.
Location/Qualifiers

FEATURES
source
1..164

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2125733"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NbHSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HPF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 69 a 19 c 19 g 57 t
ORIGIN

Query Match 5.6%; Score 18; DB 10; Length 164;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAATAATTGTAATGTT 18
|||||

Db 83 GAAATAATTGTAATGTT 66

RESULT 15
AV298638/c

LOCUS
DEFINITION AV298638 RIKEN full-length enriched, 8 days embryo Mus musculus

AV298638 177 bp mRNA EST 10-NOV-1999
cDNA clone 5730459P20 3' similar to AF045253 Mus musculus clone 4b
tousled-like kinase (mttl) mRNA, alternatively spliced, mRNA
sequence.
AV298638
AV298638.1 GI:6331017
EST.
house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 177)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,K., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno.H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .177
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5730459p20"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGATTCGAGTTAATTAATTCCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATTAATTAATTCCTCTTTTTTTTTTTNN 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT

ORIGIN

39 a 43 c 32 g 63 t

Query Match

Best Local Similarity 100.0%; Pred. No. 87;

5.6%; Score 18; DB 10; Length 177;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 167 TTCAAAATTGTGAAGAAGT 184
|||||
Db 107 TTCAAAATTGTGAAGAAGT 90
|||||

Search completed: March 30, 2002, 06:44:20
Job time: 5283 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:19:47 ; Search time 89.47 Seconds
(without alignments)
810.025 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAAATAATGTAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_NA:
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. 1s the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	10.3	33	4	US-09-081-345-9
2	33	10.3	33	4	US-09-081-345-9
3	18	5.6	1398	2	US-08-896-320-2
4	16	5.0	618	4	US-09-328-111-830
5	16	5.0	634	4	US-09-328-111-825
6	16	5.0	965	2	US-09-141-135-1
7	16	5.0	2040	3	US-08-581-148C-21
8	16	5.0	2445	4	US-09-298-367B-9
9	16	5.0	2539	3	US-08-581-148C-20
10	16	5.0	2829	3	US-08-911-321-1
11	16	5.0	2844	3	US-08-815-809-2
12	16	5.0	2856	2	US-08-816-155B-4
13	16	5.0	2856	3	US-09-079-587-4
14	16	5.0	3060	1	US-08-286-305A-6
15	16	5.0	3060	2	US-08-441-104A-6
16	16	5.0	3060	2	US-08-440-816A-6
17	16	5.0	3060	4	US-09-417-381A-6
18	16	5.0	3194	2	US-08-359-705B-1
19	16	5.0	3194	2	US-08-286-846A-1
20	16	5.0	3194	2	US-08-457-880A-1
21	16	5.0	3194	3	US-08-444-622A-1
22	16	5.0	3194	3	US-08-942-562-1
23	16	5.0	3194	3	US-09-156-923-1
24	16	5.0	3339	3	US-08-911-321-5
25	16	5.0	3707	1	US-08-271-454-1
26	16	5.0	3707	5	PCT-US95-08180-1
27	16	5.0	4524	2	US-08-845-998-7

28	16	5.0	4524	3	US-09-206-537-7	Sequence 7, Appli
29	16	5.0	4524	4	US-09-430-854-7	Sequence 7, Appli
30	16	5.0	4833	4	US-09-066-047-1	Sequence 1, Appli
31	16	5.0	6628	3	US-08-815-809-3	Sequence 3, Appli
32	16	5.0	6649	2	US-08-816-155B-5	Sequence 5, Appli
33	16	5.0	6649	3	US-09-079-587-5	Sequence 5, Appli
34	16	5.0	8501	4	US-09-298-367B-6	Sequence 5, Appli
35	15	4.7	34	4	US-09-081-345-7	Sequence 7, Appli
36	15	4.7	34	4	US-09-081-345-7	Sequence 7, Appli
37	15	4.7	34	4	US-09-081-345-7	Sequence 7, Appli
38	15	4.7	486	2	US-08-506-864A-5	Sequence 5, Appli
39	15	4.7	486	2	US-08-851-968-5	Sequence 5, Appli
40	15	4.7	536	2	US-08-341-568-1	Sequence 1, Appli
41	15	4.7	536	2	US-08-911-020-1	Sequence 1, Appli
42	15	4.7	548	1	US-08-469-667-6	Sequence 6, Appli
43	15	4.7	548	5	PCT-US95-07289-6	Sequence 6, Appli
44	15	4.7	569	1	US-08-620-467A-6	Sequence 6, Appli
45	15	4.7	569	1	US-08-348-572-6	Sequence 6, Appli
	15	4.7	569	3	US-09-041-090B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-081-345-9
; Sequence 9, Application US/09081345
; Patent No. 6228641
; GENERAL INFORMATION:
; APPLICANT: Bahija Jallal
; APPLICANT: Gregory D. Plovman
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
; TITLE OF INVENTION: PTP04 RELATED DISORDERS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Fastseq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,345
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/047,222
; FILING DATE: May 20, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 234/253
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-081-345-9

Query Match 10.3%; Score 33; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 2e-08;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CCGTTATGTGAGGAGCCACATTACAGGACC 311
|||||
Db 1 CCGTTATGTGAGGAGCCACATTACAGGACC 33

RESULT 2

US-09-081-345-10/c
: Sequence 10, Application US/09081345
: Patent No. 6228641
: GENERAL INFORMATION:
: APPLICANT: Bahija Jallal
: APPLICANT: Gregory D. Plowman
: TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
: TITLE OF INVENTION: PTP04 RELATED DISORDERS
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: FastSeq for Windows 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/081,345
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/047,222
: FILING DATE: May 20, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 234/253
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 33 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-09-081-345-10

Query Match 10.3%; Score 33; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CCGTTATGTGAGGAGCCACATTACAGGACC 311
|||||
Db 33 CCGTTATGTGAGGAGCCACATTACAGGACC 1

RESULT 3

US-08-896-320-2
: Sequence 2, Application US/08896320
: Patent No. 5871971
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Lal, Preeti
: APPLICANT: Corley, Neil C.
: APPLICANT: Shah, Purvi

: TITLE OF INVENTION: HUMAN DEVELOPMENTALLY REGULATED
: TITLE OF INVENTION: GTP-BINDING PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/896,320
: FILING DATE: Herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0344 US
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1398 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: CARDNOT01
: CLONE: 281964
US-08-896-320-2

Query Match 5.6%; Score 18; DB 2; Length 1398;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 TTCAATTGTGAGAGGT 184
|||||
Db 1148 TTCAATTGTGAGAGGT 1165

RESULT 4

US-09-328-111-830
: Sequence 830, Application US/09328111
: Patent No. 6262333
: GENERAL INFORMATION:
: APPLICANT: Endege, Wilson O.
: APPLICANT: Steinmann, Kathleen E.
: APPLICANT: Astle, Jon H.
: APPLICANT: Burgess, Christopher C.
: APPLICANT: Bushnell, Steven E.
: APPLICANT: Carroll III, Eddie
: APPLICANT: Catino, Theodore J.
: APPLICANT: Derti, Adnan
: APPLICANT: Ford, Donna M.
: APPLICANT: Lewis, Marcia E.
: APPLICANT: Monahan, John E.
: APPLICANT: Schlegel, Robert
: TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
: TITLE OF INVENTION: PRODUCTS
: FILE REFERENCE: CCD-257 (US)
: CURRENT APPLICATION NUMBER: US/09/328,111
: CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 830
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(618)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-830

Query Match 5.0%; Score 16; DB 4; Length 618;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 196 AGTCACTCTGTAAAC 211
Db 517 agtcaactctgtaaac 532

RESULT 5

US-09-328-111-825/c
; Sequence 825, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astie, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 825
; LENGTH: 634
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(634)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-825

Query Match 5.0%; Score 16; DB 4; Length 634;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 186 CACAGGAAGAGTCAC 201
Db 160 CACAGGAAGAGTCAC 145

RESULT 6

US-09-141-135-1/c
; Sequence 1, Application US/09141135

; Patent No. 5981729
; GENERAL INFORMATION:
; APPLICANT: CHUN, Jong Yoon
; APPLICANT: LEE, Yong Hun
; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisi
; FILE OF INVENTION: Acid Isolated from Arabidopsis thaliana
; FILE REFERENCE: 1942/31
; CURRENT APPLICATION NUMBER: US/09/141,135
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 1
; LENGTH: 965
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-141-135-1

Query Match 5.0%; Score 16; DB 2; Length 965;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 ATTCAAATGTGAAG 180
Db 377 ATTCAAATGTGAAG 362

RESULT 7

US-08-581-148C-21
; Sequence 21, Application US/08581148C
; Patent No. 6060844
; GENERAL INFORMATION:
; APPLICANT: Schnable, Patrick S.
; APPLICANT: Robertson, Donald S.
; APPLICANT: Hansen, Joel D.
; APPLICANT: Nikolau, Basil J.
; APPLICANT: Xu, Xiaojie
; APPLICANT: Xia, Yiji
; TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID
; TITLE OF INVENTION: GENES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voigt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/581,148C
; FILING DATE: 29-DEC-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Larcher, Carol
; REGISTRATION NUMBER: 35243
; REFERENCE/DOCKET NUMBER: 71380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS

LOCATION: 270..716
FEATURE: CDS
NAME/KEY: 1060..1875
LOCATION: 1060..1875
US-08-581-148C-21

Query Match 5.0%; Score 16; DB 3; Length 2040;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 TTTCAGATACTCAGT 148
|||||
DB 1054 TTTCAGATACTCAGT 1069

RESULT 8

US-09-298-367B-9
Sequence 9, Application US/09298367B
Patent No. 6180112
GENERAL INFORMATION:
APPLICANT: Highlander, Sarah K.
APPLICANT: Federova, Natalie D.
TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA VACCINE
FILE REFERENCE: BCM-03728
CURRENT APPLICATION NUMBER: US/09/298,367B
CURRENT FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: 08/834,455
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 2445
TYPE: DNA
ORGANISM: Pasteurella haemolytica
US-09-298-367B-9

Query Match 5.0%; Score 16; DB 4; Length 2445;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 ATGGCACTCTGCCTC 254
|||||
DB 1772 atggcactctgcctc 1787

RESULT 9

US-08-581-148C-20
Sequence 20, Application US/08581148C
Patent No. 6060644
GENERAL INFORMATION:
APPLICANT: Schnable, Patrick S.
APPLICANT: Robertson, Donald S.
APPLICANT: Hansen, Joel D.
APPLICANT: Nikolau, Basil J.
APPLICANT: Xia, Yiji
APPLICANT: Xu, Xiaojie
TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Volt & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/581,148C
FILING DATE: 29-DEC-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Larcher, Carol
REGISTRATION NUMBER: 35243
REFERENCE/DOCKET NUMBER: 71380
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-581-148C-20

Query Match 5.0%; Score 16; DB 3; Length 2539;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 TTTCAGATACTCAGT 148
|||||
DB 1354 TTTCAGATACTCAGT 1369

RESULT 10

US-08-911-321-1/c
Sequence 1, Application US/08911321
Patent No. 6010703
GENERAL INFORMATION:
APPLICANT: Roger K. Maes and Stephen J. Spatz
TITLE OF INVENTION: Recombinant Poxvirus
TITLE OF INVENTION: Vaccine Against
TITLE OF INVENTION: Feline Rhinotracheitis
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/096,183
FILING DATE: July 26, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MSU 4.1-166
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 6010703e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2829
TYPE: Nucleic Acid
STRANDEDNESS: Double

TOPOLOGY: Linear
MOLECULE TYPE: DNA
DESCRIPTION: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Feline herpesvirus-1
STRAIN: 1
INDIVIDUAL ISOLATE: C-27
CELL TYPE: N/A
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD: Sequencing
OTHER INFORMATION: DNA and deduced
OTHER INFORMATION: polypeptide
US-08-911-321-1

Query Match 5.0%; Score 16; DB 3; Length 2829;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 ATTCAAAATTGTGAAG 180
|||||

DB 1235 ATTCAAAATTGTGAAG 1220

RESULT 11

US-08-815-809-2/c
Sequence 2, Application US/08815809
Patent No. 6004777
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, James
APPLICANT: GOEBEL, Scott J.
APPLICANT: COX, William I.
APPLICANT: GETTIG, Russell R.
APPLICANT: PINCUS, Steven E.
APPLICANT: PAOLETTI, ENZO
APPLICANT: JACOBS, Bertram L.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
TITLE OF INVENTION: MAKING AND USES THEREOF
CURRENT APPLICATION NUMBER: US/08/815,809
CURRENT FILING DATE: 1997-03-12
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2844
TYPE: DNA
ORGANISM: Vaccinia virus
US-08-815-809-2

Query Match 5.0%; Score 16; DB 3; Length 2844;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 ATTCAAAATTGTGAAG 180
|||||

DB 1247 ATTCAAAATTGTGAAG 1232

RESULT 12

US-08-816-155B-4/c
Sequence 4, Application US/08816155B
Patent No. 5990091
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, James
APPLICANT: COX, William I.
APPLICANT: GETTIG, Russell R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: PAOLETTI, ENZO

APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/816,155B
FILING DATE: 12-MAR-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-816-155B-4
Query Match 5.0%; Score 16; DB 2; Length 2856;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 ATTCAAAATTGTGAAG 180
|||||
DB 1259 ATTCAAAATTGTGAAG 1244
RESULT 13
US-09-079-587-4/c
Sequence 4, Application US/09079587
Patent No. 6130066
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, JAMES
APPLICANT: COX, WILLIAM I.
APPLICANT: GETTIG, RUSSELL R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: PAOLETTI, ENZO
APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,587
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/816,155
FILING DATE: 12-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-079-587-4

Query Match 5.0% Score 16; DB 3; Length 2856;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 ATTTCAAAATTGTGAAG 180
|||||
DB 1259 ATTTCAAAATTGTGAAG 1244

RESULT 14

US-08-286-305A-6
Sequence 6, Application US/08286305A
Patent No. 5766863

GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Sadick, Michael D.
APPLICANT: Shelton, David L.
APPLICANT: Wong, Wai Lee Tan
TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,305A
FILING DATE: 05-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 854C1P1
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881

TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3060 bases
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-286-305A-6

Query Match 5.0% Score 16; DB 1; Length 3060;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 GAAGGAGCCTTTGGAA 102
|||||
DB 2017 GAAGGAGCCTTTGGAA 2032

RESULT 15

US-08-441-104A-6
Sequence 6, Application US/08441104A
Patent No. 5891650

GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Sadick, Michael D.
APPLICANT: Shelton, David L.
APPLICANT: Wong, Wai Lee Tan
TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 720 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,104A
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286305
FILING DATE: 05-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 854C1P1C2
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3060 bases
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-441-104A-6

Query Match 5.0%; Score 16; DB 2; Length 3060;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 87 GAAGGAGCCTTTGGAA 102
|||||
Db 2017 GAAGGAGCCTTTGGAA 2032

Search completed: March 30, 2002, 07:10:46
Job time: 6659 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 30, 2002, 05:16:48 ; Search time 1463.6 Seconds
(without alignments)
3606.926 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAATAATGTAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table: OLIGO_NUC
Gapex 60.0 , Gapext 60.0
Searched: 1472140 seqs, 8248589755 residues

Word size : 0
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo.hum.*
- 31: em.htgo.inv.*
- 32: em.htgo.ro.*
- 33: em.htg.hum.*
- 34: em.htg.inv.*
- 35: em.htg.ro.*
- 36: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	50	15.6	3090	10	D64141	Mus musculus
2	33	10.3	33	6	ARI49921	Sequence
3	33	10.3	33	6	ARI49922	Sequence
4	28	8.8	2786	9	HS4800374	Human DNA
5	28	8.8	197389	9	AL358791	Human DNA
6	21	6.6	137584	2	AC005845	Human DNA
7	20	6.2	95241	9	HS39819	Human DNA
8	20	6.2	148269	2	AC026486	Human DNA
9	20	6.2	158312	2	AC083922	Human DNA
10	20	6.2	162249	2	AC020682	Human DNA
11	20	6.2	166973	2	AL451049	Human DNA
12	20	6.2	188026	2	AL356952	Human DNA
13	20	6.2	273403	2	AB011498	Human DNA
14	19	5.9	1254	8	AB011796	Citrus un
15	19	5.9	8935	14	AF179368	HIV-1 str
16	19	5.9	79262	8	AC006587	Arabidops
17	19	5.9	86629	9	AF305873	Human sapi
18	19	5.9	86667	9	AF230666	Human sapi
19	19	5.9	89122	9	AC007199	Human sapi
20	19	5.9	97168	9	AF230667	Human sapi
21	19	5.9	97693	9	AC002383	Human sapi
22	19	5.9	125066	9	AC074286	Human sapi
23	19	5.9	131611	9	AC005099	Human sapi
24	19	5.9	148245	2	AC016209	Human sapi
25	19	5.9	153000	2	AC010934	Human sapi
26	19	5.9	162098	2	AC019032	Human sapi
27	19	5.9	167880	9	HU091323	Human Chrom
28	19	5.9	171374	9	AC010980	Human sapi
29	19	5.9	171769	8	FL063	Sequence
30	19	5.9	179441	2	AC044879	Human sapi
31	19	5.9	179728	2	AC087820	Human sapi
32	19	5.9	180347	9	AC009415	Human sapi
33	19	5.9	183610	2	AC026452	Human sapi
34	19	5.9	198740	2	AC069434	Human sapi
35	19	5.9	211607	2	AC025778	Human sapi
36	18	5.6	332	6	AX150131	Sequence
37	18	5.6	718	9	HS4801361	Human sapi
38	18	5.6	1380	10	AF232010	Cavia sp.
39	18	5.6	1383	9	HS48005940	Human sapi
40	18	5.6	1398	6	AR035970	Sequence
41	18	5.6	1477	9	AF131829	Human sapi
42	18	5.6	1647	9	AF078103	Human sapi
43	18	5.6	1666	11	G06702	human STS
44	18	5.6	1910	8	CBU53145	Candida boi
45	18	5.6	2137	8	YSAPMP47	Candida boi

ALIGNMENTS

RESULT : 1

D64141

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

ABSTRACT

COMMENT

FEATURES

BASES

SEQUENCE

REFERENCES

ADDITIONAL

NOTES

OTHER

END

D64141 Mus musculus mRNA 3090 bp 13-FEB-1999
Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.
D64141
D64141.1 GI:2665457
protein-tyrosine-phosphatase.
Mus musculus testis cDNA to mRNA.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3090)
Ohsugi,M.
Direct Submission
Submitted (15-SEP-1995) to the DDBJ/EMBL/GenBank databases. Miho
Ohsugi, The University of Tokyo, Department of Oncology; 4-6-1
Shirokanedai, Minato-ku, Tokyo 108, Japan
(E-mail:mohsugi@ims.u-tokyo.ac.jp, Tel:03-5449-5305,
Fax:03-5449-5413)
2 (sites)

AUTHORS Ohsugi, M., Kuramochi, S., Matsuda, S. and Yamamoto, T.
TITLE Molecular cloning and characterization of a novel cytoplasmic protein-tyrosine phosphatase that is specifically expressed in spermatocytes
JOURNAL J. Biol. Chem. 272 (52), 33092-33099 (1997)
MEDLINE 98070510
FEATURES Location/Qualifiers
source 1..3090
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="testis"
218..1498
/EC_number="3.1.3.48"
/codon_start=1
/product="protein-tyrosine-phosphatase"
/protein_id="BAA23761.1"
/db_xref="GI:2665458"
/translations="MSSPRKVRGKTGRDNDDESGNSGLNLRSLSPSSOKWTPTPKI
FGNKMSENKPSHHLSFSDKYELVPELESDDTETVMDVSDRSLRNWNMSDSETA
GPSKTVSPVLSSGSRSLKSDTETSVSEKELTQLAIRPLFNSARSAMRDCLNTLQK
EELDIREFLEQMTLPDDFNSGNTLQNRKNRYRDIPLDYSTRVPLGKNKDYINAS
YIRVHNEEYFYATQGPLPETIEDFQMWLENNGNVIAMITREIECGVICKYSWP
LSLKEPLEPHFSVLETHVTOYFTVRVFIQVKGSKOCVKHLQFTKWPDPHGTPA
SADFEIKYIVYRKSHITGPLLHVCAGVGTGCTFICVDVWFSAIEKNYSFDIMNIVT
QMRKQKGMITQEQFOFYEIVLEVLNLLALY"
BASE COUNT 904 a 603 c 647 g 936 t
ORIGIN

Query Match 15.6%; Score 50; DB 10; Length 3090;
Best Local Similarity 100.0%; Pred. No. 2.9e-16;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 271 AATATGTCGCTTATGTGAGGAGGAGCCACATTACAGGACCCCTCTGTGT 320
|||||
Db 1238 AATATGTCGCTTATGTGAGGAGGAGCCACATTACAGGACCCCTCTGT 1287
|||||

RESULT 2
AR149921
LOCUS AR149921 33 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 9 from patent US 6228641.
ACCESSION AR149921
VERSION AR149921.1 GI:15114512
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 33)
AUTHORS Jallal, B. and Plozman, G.D.
TITLE Diagnosis and treatment of PTP04 related disorders
JOURNAL Patent: US 6228641-A 9 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..33
/organism="unknown"
BASE COUNT 10 a 8 c 9 g 6 t
ORIGIN

Query Match 10.3%; Score 33; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 279 CCGTTATGTGAGGAGGAGCCACATTACAGGACC 311
|||||
Db 1 CCGTTATGTGAGGAGGAGCCACATTACAGGACC 33

RESULT 3
AR149922/c
LOCUS AR149922 33 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 10 from patent US 6228641.
ACCESSION AR149922

AR149922.1 GI:15114513
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 33)
AUTHORS Jallal, B. and Plozman, G.D.
TITLE Diagnosis and treatment of PTP04 related disorders
JOURNAL Patent: US 6228641-A 10 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..33
/organism="unknown"
BASE COUNT 6 a 9 c 8 g 10 t
ORIGIN

Query Match 10.3%; Score 33; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 279 CCGTTATGTGAGGAGGAGCCACATTACAGGACC 311
|||||
Db 33 CCGTTATGTGAGGAGGAGCCACATTACAGGACC 1

RESULT 4
HSM800374
LOCUS HSM800374 2786 bp mRNA PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFp566K0524 (from clone DKFp566K0524); partial cds.
ACCESSION AL050040
VERSION AL050040.1 GI:4884281
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2786)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Ansoorge, W., Wirkner, U., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
JOURNAL Direct Submission
COMMENT Submitted (15-MAY-1999) MIPS, Am Klopferspitze 18a, D-82152 Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFp566K0524) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES Location/Qualifiers
source 1..2786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFp566K0524"
/clone_lib="566 (synonym: hfkdx2). Vector pAMP1; host XL-2blue; sites NotI + SalI"
/dev_stage="fetal"
/tissue_type="kidney"
1..1198
/gene="DKFp566K0524"
<1..1198
/gene="DKFp566K0524"
/note="strong similarity to protein-tyrosine-phosphatases"
/codon_start=2
/product="hypothetical protein"
/protein_id="CAB43248.1"
/db_xref="GI:4884282"
/db_xref="SPTREMBL:O9Y406"
/translation="GNWRGCMGCTGVGAAPQAVFENKVNSEKVLKLSLRNPHNDY
EDVFEPSGSGDPSMWARGPRRWRSEDEAAGPSQALSPLLSDTRKIVSEGL

DOLA1RPLIENHEOTAKDCLEKTRAYDIMOEFALEKLNLPGEVSGNOPS
NREKNRYRDLDPYSTRVPLGSKSDYINASYIRVNCGEFFYIATQGLSLTDDFW
QWLENNNSNVIAMITREMEGLIKCYHYWPISLKKPLEKHFVLENYQILQYFIIR
MFQVVEKSTGTSKQLOFTKPDHGTSPASADSIKYIRYARKSHLTGPPWVHCSAG
IGRTGVFLCVDVVFCAIVKDCGSFNIMDIVAQMREQRSQMVQTKQYHFCYDITVLEVR
KLLTLD"

polya_signal 2761. .2766

polya_site 2776

BASE COUNT 841 a 494 c 550 g 901 t

ORIGIN

Query Match 8.8%; Score 28; DB 9; Length 2786;

Best Local Similarity 100.0%; Pred. No. 0.0003;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 TGGCCAGACCATGGCACTCTGCTCAG 256

|||||

Db 896 TGGCCAGACCATGGCACTCTGCTCAG 923

RESULT 5

AL358791/c

LOCUS

DEFINITION Human DNA sequence from clone RP11-42B19 on chromosome 10, complete

sequence.

ACCESSION

AL358791

VERSION

AL358791.24 GI:14141350

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 197389)

Clark,S.

Direct Submission

Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerquests@sanger.ac.uk

On May 17, 2001 this sequence version replaced gi:13751420.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; SW.,

SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-42B19 is from the library RPCI-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-42B19.

Location/Qualifiers

1..197389

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-42B19"

FEATURES

source

/clone_lib="RPCI-11.1"

6..370

/note="match: GSS: Em:AQ116178"

26..180

/note="match: GSS: Em:AQ116242"

76..566

/note="match: GSS: Em:AQ726140"

1049..1114

/note="6 copies 11 mer 83% conserved"

1192..1246

/note="5 copies 11 mer 87% conserved"

1246..1660

/note="12 repeat: matches 2256. 2667 of consensus"

2964..3011

/note="LIM3e repeat: matches 575. 621 of consensus"

3170..3580

/note="MSTA repeat: matches 1. 426 of consensus"

4275..4345

/note="MIR repeat: matches 106. 181 of consensus"

5259..5374

/note="MIR repeat: matches 111. 227 of consensus"

5414..5454

/note="L2 repeat: matches 2657. 2697 of consensus"

5633..6052

/note="LTR12 repeat: matches 1. 342 of consensus"

6053..6357

/note="Alusq repeat: matches 1. 304 of consensus"

6358..6450

/note="LTR12 repeat: matches 342. 440 of consensus"

9115..9246

/note="12 copies 11 mer 65% conserved"

9122..9261

/note="35 copies 4 mer aaga 66% conserved"

11013..11192

/note="MER45 repeat: matches 1. 178 of consensus"

11231..11638

/note="match: GSS: Em:AQ210526"

11616..12203

/note="MER4B repeat: matches 1. 611 of consensus"

12204..12435

/note="L2 repeat: matches 2251. 2500 of consensus"

12452..12704

/note="MIR repeat: matches 2. 262 of consensus"

12796..12883

/note="L2 repeat: matches 2641. 2734 of consensus"

12939..13288

/note="match: GSS: Em:A2694508"

13577..13719

/note="FLAM_C repeat: matches 1. 143 of consensus"

13747..13866

/note="MER93 repeat: matches 110. 230 of consensus"

13895..14079

/note="MER93 repeat: matches 215. 397 of consensus"

14090..14220

/note="MER5B repeat: matches 25. 178 of consensus"

14657..14971

/note="AlusC repeat: matches 1. 309 of consensus"

15219..16945

/note="L1MD repeat: matches 1905. 3500 of consensus"

16946..17256

/note="AluY repeat: matches 1. 311 of consensus"

17257..18226

/note="L1MD repeat: matches 977. 1905 of consensus"

18902..19076

/note="L1MD repeat: matches 15. 174 of consensus"

20392..20519

/note="MLTII repeat: matches 6. 129 of consensus"

20677..20808

/note="FLAM_C repeat: matches 2. 133 of consensus"

20833..20868

/note="MLTII repeat: matches 308. 346 of consensus"

20869..21214

/note="MLTIAI repeat: matches 1. 365 of consensus"

repeat_region 21217. .21621
/note="MLr1A1-internal repeat: matches 927. .1333 of
consensus"
repeat_region 22088. .22592
/note="MLr1A1-internal repeat: matches 160. .712 of
consensus"
repeat_region 22617. .22732
/note="L1ME3 repeat: matches 6022. .6145 of consensus"
repeat_region 22795. .22911
/note="MLr1A1-internal repeat: matches 5. .121 of
consensus"
repeat_region 22917. .23273
/note="MLr1A1 repeat: matches 1. .365 of consensus"
repeat_region 23274. .23344
/note="MLr1I repeat: matches 339. .410 of consensus"
repeat_region 24336. .24385
/note="MLr1F repeat: matches 268. .317 of consensus"
repeat_region 25059. .26364
/note="L1M4 repeat: matches 1064. .2552 of consensus"
repeat_region 26398. .26843
/note="L1M4 repeat: matches 2651. .3102 of consensus"
repeat_region 26879. .27248
/note="L1ME3 repeat: matches 5551. .5956 of consensus"
repeat_region 27255. .27500
/note="MLr1F repeat: matches 304. .568 of consensus"
repeat_region 27667. .27736
/note="MLr1F repeat: matches 3. .74 of consensus"
repeat_region 27759. .27927
/note="L1ME3 repeat: matches 5992. .6162 of consensus"
repeat_region 30024. .30310
/note="L1ME repeat: matches 1151. .1435 of consensus"
repeat_region 31602. .31678
/note="MIR repeat: matches 79. .155 of consensus"
misc_feature 31627. .31927
/note="match: STS: Em:HSJ10C10"
repeat_region 31786. .31900
/note="MIR repeat: matches 89. .203 of consensus"
repeat_region 32440. .32738
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 32827. .32870
/note="22 copies 2 mer tt 75% conserved"
repeat_region 34892. .35012
/note="L2 repeat: matches 2584. .2710 of consensus"
repeat_region 36225. .36549
/note="MSRD repeat: matches 21. .394 of consensus"
repeat_region 36732. .36915
/note="MIR repeat: matches 36. .234 of consensus"
repeat_region 37945. .38060
/note="MIR repeat: matches 48. .170 of consensus"
repeat_region 38422. .38524
/note="MER5A repeat: matches 9. .116 of consensus"
repeat_region 38641. .38785
/note="MIR repeat: matches 34. .192 of consensus"
repeat_region 39515. .39846
/note="L1ME3 repeat: matches 5845. .6182 of consensus"
repeat_region 40361. .41030
/note="MLr1H repeat: matches 72. .547 of consensus"
repeat_region 41704. .41789
/note="MIR repeat: matches 84. .166 of consensus"
repeat_region 41923. .42236
/note="L1ME1 repeat: matches 5808. .6163 of consensus"
repeat_region 45114. .45484
/note="MLr1A1 repeat: matches 1. .365 of consensus"
repeat_region 46495. .46591
/note="MLr1J repeat: matches 43. .140 of consensus"
repeat_region 47474. .47516
/note="L2 repeat: matches 2658. .2700 of consensus"
repeat_region 48316. .48775
/note="L1M4 repeat: matches 6443. .6923 of consensus"
repeat_region 49375. .49685
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 49686. .49720
/note="Lr12 repeat: matches 445. .479 of consensus"

repeat_region 49691. .49724
/note="PTR5 repeat: matches 173. .206 of consensus"
repeat_region 49725. .50193
/note="Lr12 repeat: matches 19. .1523 of consensus"
Query Match 8.8%; Score 28; DB 9; Length 197389;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 229 TGGCAGACCATGCGACTCTCGCCTCAG 256
Db 158585 TGGCAGACCATGCGACTCTCGCCTCAG 158558
RESULT 16 6
AC005845/c
LOCUS
DEFINITION
Homo sapiens chromosome 12 clone RP3-454B23, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
AC005845
AC005845.8 GI:14787165
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy- Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Poster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., King,L., Korvah,J.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
Oguri,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,A.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R.,
Thoms,S., Usmani,K., Vasquez,L., Viera,I., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 137584)
Worley,K.C.
Direct Submission
Submitted (22-OCT-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 17, 2001 this sequence version replaced gl:14625075.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: ZK
Center clone name: RP3-454B23
----- Summary Statistics
Sequencing vector: Plasmid; M7789
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 139066 bases at least Q40
Consensus quality: 142794 bases at least Q30
Consensus quality: 144767 bases at least Q20
Estimated insert size: 137687; sum-of-contrigs estimation
Quality coverage: 10.8x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 58304: contig of 58304 bp in length
* 58305 58404: gap of unknown length
* 58405 97131: contig of 38727 bp in length
* 97132 97231: gap of unknown length
* 97232 111131: contig of 13900 bp in length
* 111132 111231: gap of unknown length
* 111232 126332: contig of 15101 bp in length
* 126333 126432: gap of unknown length
* 126433 137584: contig of 11152 bp in length.
* 137584 Location/Qualifiers

FEATURES
source

1. 137584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP3-454B23"
BASE COUNT 36835 a 33464 c 33266 g 33618 t 401 others
ORIGIN

Query Match 6.6%; Score 21; DB 2; Length 137584;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 TTCAACACTTTTCATGTCCTT 123
|||||
Db 5514 TTCAACACTTTTCATGTCCTT 5494

RESULT 7

HS39819 95241 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from clone 39819 on chromosome 1p34.3-36.13
DEFINITION Contains acute myeloid leukaemia (AML1) gene, EST, GSS, CpG island,
complete sequence.
ACCESSION AL023096
VERSION AL023096.1 GI:3900882
KEYWORDS HTG; acute myeloid leukaemia; AML1; CpG island; runt domain;
transcription factor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

JOURNAL

COMMENT

1 (bases 1 to 95241)
Grafham,D.
Direct Submission
Submitted (04-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Nov 21, 1998 this sequence version replaced gl:3114715.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 39819. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contrigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
39819 is from the library RPC13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>.

FEATURES
source

1. 95241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p34.3-36.13"
/clone="RP3-39819"
/clone_lib="RPCI-3"
1. 242
repeat_region
555. 634
repeat_region
849. 1014
repeat_region
1022. 1249
repeat_region
/note="AluX repeat: matches 59. 302 of consensus"
/note="MIR repeat: matches 69. 147 of consensus"
/note="AluSx repeat: matches 1. 168 of consensus"
/note="AluJo repeat: matches 85. 301 of consensus"
complement(1624. 1743)
/note="MIR repeat: matches 151. 33 of consensus"
1888. 2058
complement(2098. 2239)
/note="MIR repeat: matches 21. 195 of consensus"
/note="MIR2 repeat: matches 146. 4 of consensus"
3109. 3175
/note="MIR repeat: matches 73. 139 of consensus"
3924. 4206
/note="AluSg repeat: matches 1. 298 of consensus"
4270. 4481
/note="MER42c repeat: matches 1248. 1481 of consensus"
complement(4490. 4789)
/note="AluSx repeat: matches 301. 2 of consensus"
complement(4790. 4915)
/note="AluSx repeat: matches 136. 1 of consensus"
complement(5362. 5662)
/note="AluSg repeat: matches 303. 1 of consensus"
5918. 6029
/note="MIR repeat: matches 35. 147 of consensus"
complement(7365. 7572)
/note="MIR repeat: matches 240. 14 of consensus"
complement(8168. 8279)
/note="MIR repeat: matches 192. 80 of consensus"
8835. 9131
/note="AluSx repeat: matches 3. 302 of consensus"
9426. 9721
/note="AluJo repeat: matches 1. 300 of consensus"
complement(10746. 11046)
/note="AluSx repeat: matches 302. 1 of consensus"
11707. 11857
repeat_region

```

repeat_region /note="MIR repeat: matches 34. .179 of consensus"
11928. .12036
/note="MER5A repeat: matches 2. .108 of consensus"
repeat_region complement(12047. .12348)
/note="AluX repeat: matches 302. .1 of consensus"
repeat_region complement(12368. .12685)
/note="LMB7 repeat: matches 907. .574 of consensus"
repeat_region complement(12783. .13172)
/note="LMB1 repeat: matches 1063. .672 of consensus"
repeat_region complement(13175. .13475)
/note="MSTB repeat: matches 392. .82 of consensus"
repeat_region complement(13478. .13776)
/note="AluJ repeat: matches 300. .1 of consensus"
repeat_region complement(13777. .13869)
/note="MSTC repeat: matches 97. .1 of consensus"
repeat_region complement(13870. .15565)
/note="MST-INTERNAL repeat: matches 1651. .1 of consensus"
repeat_region complement(15561. .15601)
/note="MSTB repeat: matches 382. .342 of consensus"
repeat_region complement(15679. .15986)
/note="MSTB repeat: matches 322. .1 of consensus"
repeat_region complement(15987. .16660)
/note="LMB1 repeat: matches 676. .5 of consensus"
repeat_region complement(16516. .19215)
/note="L1 repeat: matches 5390. .2714 of consensus"
repeat_region complement(19218. .19320)
/note="LMB5 repeat: matches 573. .471 of consensus"
repeat_region complement(19327. .19977)
/note="LMB6 repeat: matches 880. .224 of consensus"
repeat_region complement(19345. .20266)
/note="LMB3 repeat: matches 878. .1 of consensus"
repeat_region complement(20115. .20307)
/note="L1 repeat: matches 5390. .5201 of consensus"
repeat_region complement(20337. .20638)
/note="AluJ repeat: matches 302. .1 of consensus"
repeat_region complement(20661. .20882)
/note="AluJ repeat: matches 298. .79 of consensus"
repeat_region complement(20888. .21175)
/note="AluSp repeat: matches 300. .3 of consensus"
repeat_region complement(21939. .22342)
/note="MLT1C repeat: matches 466. .53 of consensus"
repeat_region 22549. .22632
/note="42 copies 2 mer ca 86% conserved"
repeat_region 22552. .22635
/note="6 copies 14 mer 85% conserved"
repeat_region 22698. .23001
/note="AluSp repeat: matches 1. .303 of consensus"
prim_transcript complement(23021. .23185)
/note="match: 5' EST R39234 clone 60330"
repeat_region 23819. .23854
/note="9 copies 4 mer ttat 100% conserved"
repeat_region complement(23857. .24307)
/note="LIP9 repeat: matches 910. .454 of consensus"
repeat_region complement(24336. .24754)
/note="MLT2-internal repeat: matches 3425. .2966 of consensus"
repeat_region complement(24774. .25148)
/note="MLT1C repeat: matches 414. .33 of consensus"
repeat_region complement(25194. .25333)
/note="MLT2-internal repeat: matches 2930. .2792 of consensus"
repeat_region complement(25338. .25639)
/note="AluX repeat: matches 302. .1 of consensus"
repeat_region complement(26193. .26543)
/note="THE1B repeat: matches 364. .1 of consensus"
repeat_region complement(26544. .27915)
/note="THE1B-INTERNAL repeat: matches 1580. .207 of consensus"
repeat_region 27868. .27939
/note="THE1B repeat: matches 224. .295 of consensus"
repeat_region complement(27972. .28199)
/note="MLT2-internal repeat: matches 2626. .2397 of consensus"

repeat_region 28319. .28452
/note="MIR2 repeat: matches 1. .145 of consensus"
repeat_region complement(29645. .29945)
/note="AluX repeat: matches 301. .1 of consensus"
repeat_region 29946. .30076
/note="MIR repeat: matches 114. .262 of consensus"
repeat_region 30607. .30881
/note="AluJ repeat: matches 3. .299 of consensus"
repeat_region 34070. .34107
/note="MIR repeat: matches 108. .145 of consensus"
mRNA complement(join(34216. .37315,41908. .42066,53889. .53993,
62223. .62379,64236. .64927))
/gene="AML1"
/note="match: 235278"
/pseudo
/evidence-not_experimental
gene complement(34216. .64927)
/gene="AML1"
repeat_region 36061. .36102
/note="21 copies 2 mer tt 91% conserved"
CDS complement(join(37172. .37716,42309. .42467,54290. .54394,
62624. .62780,64637. .64918))
/gene="AML1"
/note="match: 235278"
/codon_start=1
/evidence-not_experimental
/product="AML1 (acute myeloid leukaemia)"
/protein_id="CAA18856.1"
/db_xref="GI:3900883"
/db_xref="SPTREMBL:Q13761"
/translation="MRIPVDPSTSRFTPPSPAPCGGGGKMGKSGALSAQAAYGP
GGHARPEVRMVDVLADHAGELVTRSPNPLCSVLPSHWKCNKTLPAFKVALGDVP
DETQVTVMAKNDENYSAEELRNASAVKMQVARENDLRFVGRSGKSFLLITVFNTP
TOVATYHAIKVTVDGPRPRHRQKLEQTKPDPDFGDLRLMRVTPSTSPRGS
LSTTSHFSSOPOTIOGTSELNPFSDPROFDRSFPTLPTLTESRFPDPRMHYPGMSA
APPYSATPSSTSISSLVAGMPATSRFHHTYLPPEYGPQNGSPQANPSPYHLVY
GRSSGYSQPSWAGSSGGDRSPTRMLASCTSSAASVAAGNLMNPSLGGSGDGEADG
SHSNPTALSTPGMDVAWRPY"
repeat_region complement(38672. .38726)
/note="MIR repeat: matches 133. .80 of consensus"
repeat_region 38810. .39019
/note="MIR repeat: matches 2. .243 of consensus"

Query Match 6.2%; Score 20; DB 9; Length 95241;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GTAGATTTTTCATCAATA 275
|||||
Db 17969 GTAGATTTTTCATCAATA 17988

RESULT 8
LOCUS AC026486 148269 bp DNA HTG 05-MAY-2001
DEFINITION Homo sapiens clone RP11-24K7, WORKING DRAFT SEQUENCE, 18 unordered
pieces.
ACCESSION AC026486
VERSION AC026486.2 GI:9954792
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148269)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-24K7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148269)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

```

Coddy, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dominguez, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardy, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKenna, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Olivier, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 31, 2000 this sequence version replaced GI:7280316.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4489

Center clone name: 24_K_7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 140190 bases at least Q40

Consensus quality: 144274 bases at least Q30

Consensus quality: 145694 bases at least Q20

Insert size: 145000; agarose-gel

Insert size: 146569; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-gel

Quality coverage: 4.5 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1564: contig of 1564 bp in length
1565 1664: gap of 100 bp
1665 20700: contig of 19036 bp in length
20701 20800: gap of 100 bp
20801 21849: contig of 1049 bp in length
21850 21949: gap of 100 bp
21950 25604: contig of 3655 bp in length
25605 25704: gap of 100 bp
25705 29764: contig of 4060 bp in length
29765 29864: gap of 100 bp
29865 35579: contig of 5715 bp in length
35580 35679: gap of 100 bp
35680 41577: contig of 5898 bp in length
41578 41677: gap of 100 bp
41678 47413: contig of 5736 bp in length
47414 47513: gap of 100 bp
47514 54860: contig of 7347 bp in length
54861 54960: gap of 100 bp
54961 61589: contig of 6629 bp in length
61590 61689: gap of 100 bp
61690 68215: contig of 6526 bp in length

* 58216 68315: gap of 100 bp
* 68316 78737: contig of 10422 bp in length
* 78738 78837: gap of 100 bp
* 78838 89917: contig of 11080 bp in length
* 89918 90017: gap of 100 bp
* 90018 100805: contig of 10788 bp in length
* 100806 100905: gap of 100 bp
* 100906 114689: contig of 13784 bp in length
* 114690 114789: gap of 100 bp
* 114790 132632: contig of 17843 bp in length
* 132633 132732: gap of 100 bp
* 132733 147978: contig of 15246 bp in length
* 147979 148078: gap of 100 bp in length.
* 148079 148269: contig of 191 bp in length.

FEATURES

source

1. 148269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-24K7"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. 1564
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

misc_feature

1665..20700
/note="assembly_fragment"

misc_feature

20801..21849
/note="assembly_fragment"

misc_feature

21950..25604
/note="assembly_fragment"

misc_feature

25705..29764
/note="assembly_fragment"

misc_feature

29865..35579
/note="assembly_fragment"

misc_feature

35680..41577
/note="assembly_fragment"

misc_feature

41678..47413
/note="assembly_fragment"

misc_feature

47514..54860
/note="assembly_fragment"

misc_feature

54961..61589
/note="assembly_fragment"

misc_feature

61690..68215
/note="assembly_fragment"

misc_feature

68316..78737
/note="assembly_fragment"

misc_feature

78838..89917
/note="assembly_fragment"

misc_feature

90018..100805
/note="assembly_fragment"

misc_feature

100906..114689
/note="assembly_fragment"

misc_feature

114790..132632
/note="assembly_fragment"

misc_feature

132733..147978
/note="assembly_fragment"

misc_feature

148079..148269
/note="assembly_fragment"
clone_end:T7
vector_side:right"

BASE COUNT 46902 a 26848 c 26913 g 45901 t 1705 others
ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 148269;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 TGAAGGAGCCTTTGGAATTC 105
|||||
Db 48091 TGAAGGAGCCTTTGGAATTC 48110

```

RESULT 9
AC083922/c
LOCUS
DEFINITION
Homo sapiens chromosome 8 clone RP11-63H17 map 8, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
AC083922
AC083922.2 GI:13123241
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 158312)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Bouckhagter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-pierre,N., Hagoos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRoque,K.,
Lamardes,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:10697459.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L11258
Center clone name: 63_H_17
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157158 bases at least Q40
Consensus quality: 157649 bases at least Q30
Consensus quality: 157821 bases at least Q20
Insert size: 157912; sum-of-contigs
Quality coverage: 8.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20482: contig of 20482 bp in length
* 20483 20582: gap of 100 bp
* 20583 34457: contig of 13875 bp in length

```

```

* 34458 34557: gap of 100 bp
* 34558 51848: contig of 17291 bp in length
* 51849 51948: gap of 100 bp
* 51949 100652: contig of 48704 bp in length
* 100653 100752: gap of 100 bp
* 100753 158312: contig of 57560 bp in length.
FEATURES
source
1. .158312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-63H17"
/clone_lib="RPCI-11 Human Male BAC"
1. .20482
/notes="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
20583..34457
/notes="assembly_fragment"
34558..51848
/notes="assembly_fragment"
51949..100652
/notes="assembly_fragment"
100753..158312
/notes="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 46458 a 2936l c 29496 g 52597 t 400 others
ORIGIN
Query Match 6.2%; Score 20; DB 2; Length 158312;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 204 TGTAAACACTTGCAGTTCA 223
|||||
Db 113950 TGTAAACACTTGCAGTTCA 113931
RESULT 10
AC020682
LOCUS
DEFINITION
Homo sapiens clone RP11-24G18, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
AC020682
VERSION
AC020682.3 GI:9280776
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 162249)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 162249)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Bouckhagter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagoos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

```

Tirrelli, A., Vassiliiev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 19, 2000 this sequence version replaced gi:6731246.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4453
 Center clone name: 24_G.18
 ----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 155671 bases at least Q40
 Consensus quality: 158736 bases at least Q30
 Consensus quality: 159945 bases at least Q20
 Insert size: 162000; agarose-fp
 Insert size: 161049; sum-of-contigs
 Quality coverage: 5.1 in Q20 bases; agarose-fp
 Quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1597: contig of 1597 bp in length
 * 1598 1697: gap of 100 bp
 * 1698 4621: contig of 2924 bp in length
 * 4622 4721: gap of 100 bp
 * 4722 8497: contig of 3776 bp in length
 * 8498 8597: gap of 100 bp
 * 8598 14869: contig of 6272 bp in length
 * 14870 14969: gap of 100 bp
 * 14970 20957: contig of 5988 bp in length
 * 20958 21057: gap of 100 bp
 * 21058 28867: contig of 7810 bp in length
 * 28868 28967: gap of 100 bp
 * 28968 39639: contig of 10672 bp in length
 * 39640 39739: gap of 100 bp
 * 39740 48843: contig of 9104 bp in length
 * 48844 48943: gap of 100 bp
 * 48944 63260: contig of 14317 bp in length
 * 63261 63360: gap of 100 bp
 * 63361 82571: contig of 19211 bp in length
 * 82572 82671: gap of 100 bp
 * 82672 100593: contig of 17922 bp in length
 * 100594 100693: gap of 100 bp
 * 100694 122394: contig of 21701 bp in length
 * 122395 122494: gap of 100 bp
 * 122495 162249: contig of 39755 bp in length.

FEATURES

source

1. .162249
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-24G18"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. .1597

/note="assembly_fragment"

misc_feature

1698. .4621

/note="assembly_fragment"

misc_feature

4722. .8497

misc_feature
 8598. .14869
 /note="assembly_fragment"
 misc_feature
 14970. .20957
 /note="assembly_fragment"
 misc_feature
 21058. .28867
 /note="assembly_fragment"
 misc_feature
 28968. .39639
 /note="assembly_fragment"
 misc_feature
 39740. .48843
 /note="assembly_fragment"
 misc_feature
 48944. .63260
 /note="assembly_fragment"
 misc_feature
 63361. .82571
 /note="assembly_fragment"
 misc_feature
 82672. .100593
 /note="assembly_fragment"
 misc_feature
 100694. .122394
 /note="assembly_fragment"
 misc_feature
 122495. .162249
 /note="assembly_fragment"
 misc_feature
 162249. .39755
 /note="assembly_fragment"
 misc_feature
 39755. .49515
 /note="assembly_fragment"

BASE COUNT 52573 a 28799 c 30155 g 49515 t 1207 others
 ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 162249;
 Best local Similarity 100.0%; Pred. No. 7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TGAAGGAGCCTTTGGATTC 105
 |||||
 DB 59867 TGAAGGAGCCTTTGGATTC 59886

RESULT 11
 AL451049

LOCUS
 DEFINITION Human DNA sequence from clone RP11-63A2 on chromosome 10, complete sequence.
 ACCESSION AL451049
 VERSION AL451049.11 GI:14018294
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On May 11, 2001 this sequence version replaced gi:11995192.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP

database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
Rp11-63A2 is from the library RPCI-11.1 constructed by the group of
Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6

This sequence is the entire insert of clone Rp11-63A2 The true
right end of clone Rp11-809M12 is at 49921 in this sequence.

FEATURES

Location/Qualifiers

1..16973

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="Rp11-63A2"

/clone_lib="RPCI-11.1"

3..805

/note="L1MC5 repeat: matches 7127..7913 of consensus"

1250..1679

/note="MLR2B repeat: matches 1..444 of consensus"

1682..4635

/note="HERVL repeat: matches 2635..5652 of consensus"

4636..4874

/note="AluSq repeat: matches 67..297 of consensus"

4875..7529

/note="HERVL repeat: matches 1..2635 of consensus"

7530..7577

/note="MLR2A repeat: matches 406..453 of consensus"

7594..8007

/note="MLR2A repeat: matches 150..450 of consensus"

8233..8517

/note="MLR1-INTERNAL repeat: matches 932..1211 of
consensus"

8558..8854

/note="L1MC/D repeat: matches 5313..5604 of consensus"

8913..9043

/note="L1MC/D repeat: matches 5604..5735 of consensus"

9108..9312

/note="L1MC5 repeat: matches 7273..7496 of consensus"

9469..9845

/note="L1MC5 repeat: matches 7507..7895 of consensus"

10014..10105

/note="46 copies 2 mer aa 63% conserved"

10601..10716

/note="L1M4 repeat: matches 2695..2818 of consensus"

10829..11016

/note="L1M4 repeat: matches 2957..3146 of consensus"

11333..11390

/note="MLR2D repeat: matches 1..58 of consensus"

11411..11516

/note="L1M4 repeat: matches 2957..3062 of consensus"

11612..11968

/note="MLR2D repeat: matches 70..435 of consensus"

11986..12049

/note="32 copies 2 mer aa 67% conserved"

12683..12935

/note="L1ME3A repeat: matches 5669..5936 of consensus"

13751..14155

/note="L1ME repeat: matches 471..848 of consensus"

14156..14460

/note="AluDb repeat: matches 1..300 of consensus"

14461..14976

/note="L1ME repeat: matches 848..1572 of consensus"

17233..17352

/note="3 copies 40 mer 75% conserved"

17242..17345

/note="26 copies 4 mer gaag 72% conserved"

17276..17359

/note="14 copies 6 mer agaaag 65% conserved"

18041..18273

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/note="MIR repeat: matches 8..246 of consensus"
19550..19633
/note="42 copies 2 mer ta 83% conserved"
19554..19633
/note="2 copies 40 mer 91% conserved"
20024..22103
/note="L1PB3 repeat: matches 4126..6150 of consensus"
26100..26169
/note="35 copies 2 mer at 77% conserved"
26101..26164
/note="16 copies 4 mer tata 79% conserved"
26106..26165
/note="5 copies 12 mer 81% conserved"
26204..27088
/note="L1MC3 repeat: matches 6716..7586 of consensus"
27760..27918
/note="MLR1J repeat: matches 201..346 of consensus"
28811..28944
/note="MIR repeat: matches 12..149 of consensus"
29062..29149
/note="MLR1D repeat: matches 2..89 of consensus"
30829..30884
/note="MIR repeat: matches 76..131 of consensus"
33710..33949
/note="MIR repeat: matches 13..262 of consensus"
34018..34106
/note="L2 repeat: matches 2655..2747 of consensus"
34360..34452
/note="MIR repeat: matches 50..144 of consensus"
35853..35789
/note="MIR repeat: matches 3..137 of consensus"
36130..36201
/note="12 copies 6 mer atat 69% conserved"
36131..36202
/note="36 copies 2 mer ta 70% conserved"
37945..37968
/note="12 copies 2 mer tc 100% conserved"
38866..39253
/note="L1MA3 repeat: matches 5912..6304 of consensus"
39256..39337
/note="THE1B repeat: matches 1..82 of consensus"
39340..41088
/note="L1M1 repeat: matches 971..2727 of consensus"
41115..41324
/note="L1MC5 repeat: matches 7701..7912 of consensus"
41359..41767
/note="L1MB8 repeat: matches 5725..6145 of consensus"
41771..41943
/note="L1MC5 repeat: matches 7603..7774 of consensus"
43197..43436
/note="MIR repeat: matches 16..260 of consensus"
43580..43882
/note="ALUSC repeat: matches 1..305 of consensus"
44598..44892
/note="AluSq repeat: matches 1..296 of consensus"
45115..45156
/note="L1R16B repeat: matches 113..160 of consensus"
49483..49604
/note="MIR repeat: matches 141..262 of consensus"
49614..49668
/note="L2 repeat: matches 2626..2682 of consensus"
50064..50218
/note="MIR repeat: matches 12..197 of consensus"
50244..50526
/note="ALUSX repeat: matches 1..286 of consensus"
51305..51539
/note="L1PAL6 repeat: matches 5913..6157 of consensus"
52439..52618
/note="MER58A repeat: matches 19..224 of consensus"
52647..52871
/note="MIR repeat: matches 2..230 of consensus"
54233..54312
/note="2 copies 40 mer 95% conserved"

```
repeat_region 55178..55369
/Note="MLTAL repeat: matches 1..180 of consensus"
repeat_region 55428..55613
/Note="MLTAL repeat: matches 181..374 of consensus"
repeat_region 55992..56287
/Note="ALUSX repeat: matches 3..307 of consensus"
repeat_region 57708..57735
/Note="14 copies 2 mer aa 100% conserved"
repeat_region 59512..59581
/Note="35 copies 2 mer tc 92% conserved"
repeat_region 59514..59581
/Note="17 copies 4 mer tctc 94% conserved"
repeat_region 60185..60295
/Note="MIR repeat: matches 6..120 of consensus"
repeat_region 61009..61056
/Note="8 copies 6 mer tacaca 79% conserved"
repeat_region 61011..61054
/Note="11 copies 4 mer caca 81% conserved"
repeat_region 61281..62062
/Note="L2 repeat: matches 1849..2714 of consensus"
repeat_region 62072..62276

Query Match 6.2%; Score 20; DB 9; Length 166973;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TGAAGGAGCCTTTGGAAATTC 105
|||||
Db 12948 TGAAGGAGCCTTTGGAAATTC 12967

RESULT 12
AL356952
LOCUS AL356952 188026 bp DNA HTG 21-JUL-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-809M12, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL356952
VERSION AL356952.17 GI:15020905
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188026)
Direct Submission
Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970342.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA809M12
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: piasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads
Chemistry: Dye-primer Big Dye; 2% of reads
Consensus quality: 187552 bases at least Q40
Consensus quality: 187814 bases at least Q30
Consensus quality: 187920 bases at least Q20
Insert size: 188026; sum-of-contigs
Insert size: 167487; 12.8% error; agarose-fp
Quality coverage: 6.20x in Q20 bases; sum-of-contigs Quality
coverage: 6.96x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
```

```
* the accession number will be preserved.
Location/Qualifiers
1..188026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-809M12"
/clone_lib="RPC1-11.3"
/misc_feature 1..188026
/Note="assembly_fragment:01381"
BASE COUNT 61435 a 34467 c 34626 g 57498 t
ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 188026;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TGAAGGAGCCTTTGGAAATTC 105
|||||
Db 154052 TGAAGGAGCCTTTGGAAATTC 154071

RESULT 13
AC011498
LOCUS AC011498 273403 bp DNA HTG 26-JUL-2001
DEFINITION Homo sapiens clone CTB-50L17, *** SEQUENCING IN PROGRESS ***, 1
ordered pieces.
ACCESSION AC011498
VERSION AC011498.6 GI:15022009
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 273403)
Direct Submission
Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 26, 2001 this sequence version replaced gi:13699566.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 273403: contig of 273403 bp in length.
----- Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 83685
Center clone name: CRB-50L17
-----
Summary Statistics
Consensus quality: 273020 bases at least Q40
Consensus quality: 273294 bases at least Q30
Consensus quality: 273384 bases at least Q20
Estimated insert size: 230000; agarose-fp estimation
Estimated insert size: 273402; sum-of-contigs
estimation
estimation
Quality coverage: 13.18 in Q20 bases; agarose-fp
estimation
```

estimation. Quality coverage: 11.09 in Q20 bases; sum-of-contigs

```

FEATURES
  source
    1. .273403
      /location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="CTB-50L17"
BASE COUNT 64486 a 73388 c 76106 g 59423 t
ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 273403;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CCAGAGATAGAGGTGGA 51
Db 156549 CCAGAGATAGAGGTGGA 156568

RESULT .14
AB011796
LOCUS 1254 bp mRNA PLN 05-FEB-1999
DEFINITION Citrus unshiu CitFLS mRNA for flavonol synthase, complete cds.
ACCESSION AB011796
VERSION AB011796.1 GI:4126402
KEYWORDS flavonol synthase.
SOURCE Citrus unshiu (cultivar:Satsuma mandarin) cDNA to mRNA.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE 1 (bases 1 to 1254)
AUTHORS Moriguchi,T.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-1998) to the DDBJ/EMBL/GenBank databases, Takaya
Moriguchi, National Institute of Fruit Tree Science; Tsukuba
Science City, Tsukuba, Ibaraki 305-8605, Japan
(8-mail: takayafruit.affrc.go.jp, Tel:81-298-38-6452,
Fax:81-298-38-6437)
REFERENCE 2 (bases 1 to 1254)
AUTHORS Moriguchi,T.
TITLE Flavonoid biosynthesis in citrus
JOURNAL Published Only in DataBase (1999) In press
FEATURES
  source
    1. .1254
      /location/Qualifiers
        /organism="Citrus unshiu"
        /cultivar="Satsuma mandarin"
        /db_xref="taxon:55188"
    93. .1100
      /gene="CitFLS"
    93. .1100
      /gene="CitFLS"
      /codon_start=1
      /product="flavonol synthase"
      /protein_id="BAA36554.1"
      /db_xref="GI:4126403"
      /translation="MEVERVQAIASLSHNGTIPAEFIRKEQEPASTTYHGPAPELP
      TIDLDPVDRILVRSIAEASREMGIFQVFNHGIIPSDICKLQAVGKEFFELPQEEKEV
      YSRPAKDVQVGTGLQKEVGKSKWVDHLFRVWPPSSINRWPKNPPSYRAVNE
      EYAKYRVREVVDKLFTYLSLGLGVEGVLKEAAGGDDIEYMLKINYYPCPPDPLALGV
      VAHTDLSALTVLVNEVPGLOVPKDRWDIDAKYIPNALVIHIGDQIEILSNKRYKAVL
      HRTYVNDKTRMSWPVLEPPADTVVGLPQLVDDENPPRYKAKRPKDYSCYCKLNKLP
      Q"
    1254
      /polyA_site
BASE COUNT 345 a 291 c 310 g 308 t
ORIGIN

Query Match 5.9%; Score 19; DB 8; Length 1254;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 CCTCAGTAGATTTTTCAT 269
Db 1092 CCTCAGTAGATTTTTCAT 1110

RESULT .15
AF179368/c
LOCUS 8935 bp DNA VRL 01-JUL-2000
DEFINITION HIV-1 strain GR17 from Greece complete genome.
ACCESSION AF179368
VERSION AF179368.1 GI:8886632
KEYWORDS Human immunodeficiency virus type 1.
SOURCE Human immunodeficiency virus type 1.
ORGANISM Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
lentivirus group.
REFERENCE 1 (bases 1 to 8935)
AUTHORS Patskevits,D., Maglorkinis,M., Paparizos,V., Pavlakis,G.N. and
Hatzakis,A.
TITLE Molecular characterization of a recombinant HIV type 1 isolate
(A/G/E/?): unidentified regions may be derived from parental
subtype E sequences
JOURNAL AIDS Res Hum. Retroviruses 16 (9), 845-855 (2000)
MEDLINE 20331865
REFERENCE 2 (bases 1 to 8935)
AUTHORS D. Patskevits, M. Maglorkinis, V. Paparizos, G.N. Pavlakis and A.
Hatzakis.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Hygiene and Epidemiology, Athens University
Medical School, Mikrasias 75, Athens 11527, Greece
FEATURES
  source
    1. .8935
      /organism="Human immunodeficiency virus type 1"
      /strain="GR17"
      /db_xref="taxon:11676"
      /country="Greece"
      /location/Qualifiers
        107. .1594
          /gene="gag"
          /notes="corresponds to gag polyprotein"
          107. .1594
            /gene="gag"
            <1390. .4398
              /gene="pol"
              <1390. .4398
                /gene="pol"
                /notes="pol polyprotein"
                /codon_start=1
                /product="pol"
                /protein_id="AAF80532.1"
                /db_xref="GI:8886633"
                /translation="FFENLAFQOGEAREFPTEQARANSPTRELVRGDNPLPETG
                AGEGTISYFPQITLWQRPVVKVAGOLKALDGTADTVLEMSLPGRWKPKMI
                GGIGFIKVKQYEDITIEGKKAIGTVLIGTPVNIIGRNLITGCTLNPISPVD
                TVPVKLPQMDGPKYKQWPLETEKIKALTEICTEMEKEKISKIGPENYINPVFAIK
                KKDTRKRLVDFRELNRKQDFWEIQLPAGLKKKSVTLVDGDAYFSVLIDE
                SRKYTAFTIPSLNNETPGIYQYINVLPGQWGSFAIQSSMTKILEPFTQNPVEVI
                WYQIDLPDKESWTVDIOLKVLNMAQSYAGIKVKOLCKLLRGAKALTDIVLPT
                AEAELNENRELLKEPVHGVYDPSKDLIAEVOKGOLQDWTYQIYQEPKNLKTKY
                AKRRTAHTNDVQLAEVOKIALEGIVINGKIPKRLPIQKETWETWTDYQWATWIP
                ENEFVNPDLVRLKLEKEPIGAEFTFDGNAANRETLKRGAKYSELSVQITQLIKK
                EFTNPKTELEAHLALKDSGLEVINVTDSOYALGIIQAQPDKSELSVQITQLIKK
                ERVYLSWPAHRKGGNEQVDRKLVSSGIRKVLFDGIDKAPEGHERYHNSNRAMASDF
                NLPPVIAKEIVANCCKQGEAMHGVDCSPRIWOLDCTHLEGTILVAVHVASGYI
                EAVPAETGQETAVFILKAGRPVVKVHTDNGSNFTSNVAKACQWAGIQOEFQIP
                YNPQSGVSVESMKNGLKEIIGOVREOAEHLKTAQVMAVFIHNFKGGKGGGYSAGRI
                IDIIATDLQTKELQKQITIKRFTVYRDSRPIWPKAPKALLWKGGGAIIVQNSDIK
                VYPRRKAKIIRIDYGMAGDDCVAGSQDSD"
          4343. .4921
            /gene="vif"
          4343. .4921
            /gene="vif"

```



```

/gene
CDS
/translation="MENRQVMYVQVQVARMIRTWNSLVKHHMVSVKAKGWLVRHHY
ESRHPYVSSEVHIPIEDMLVITTYWGLHTGEREWHLQOGVSIERRKRYRTQVDPEL
ADQILHHYFPCFAESAIRKAILGHRVSPRCEYPAGRNKVGSLQYLALKALVTPKRTK
PPLFSVKLTFEDRNKPKQTRGRGNHTVNC"
4861..5151
/gene="vpr"
4861..5151
/gene="vpr"
/note="vpr protein"
/codon_start=1
/product="vpr"
/db_xref="GI:8886635"
/translation="MEQAPEDQPOREPYSEWMLLELLEELKHEAVRHPRPWLHGLGQ
HIYNYGDTWEGVEAIRILQQLFVHFIRGICQHSRIGIIRGRGRNGARRY"
5132..7742
/gene="tat"
Join(5132..5346,7652..7742)
/gene="tat"
/note="tat protein"
/codon_start=1
/product="tat"
/db_xref="GI:8886635"
/translation="MEPGDINIDPWNQPSQPKTACQCYKRCYHCQHCFLKKGGLG
ISYGRKKWRQRTASRSSKNHODPIPEQPLQASRNPTGPEPKKEVESKAEPAPFD"
5271..7926
/gene="rev"
Join(5271..5346,7652..7926)
/gene="rev"
/note="rev protein"
/codon_start=1
/product="rev"
/db_xref="GI:8886637"
/translation="MAGRSGSDAQLAAARIILKYOSSPYPKPAGTRQARRNORRR
WRARQNQLHSIQRLSTCLGRSEPVPLPPIERLHLCSEDCGNSGTQGVGESQI
PGESSTVLGPRTER"
5534..8068
/gene="env"
5534..8068
/gene="env"
/note="env polyprotein"
/codon_start=1
/product="env"
/db_xref="GI:8886638"
/translation="MRVKETQRNHNLRWGLMIFGMLMICAEMWVTVYVYGVVVR
DADTILFCASDARTYSTEKHNWATHSCVPTDNPRLSENVTENFMWKNVVEQM
HEDVISLWDESLKPCVLTPLCVALNCTDARDNATIDIKNCTFNITTELEDDKKNEERL
FYRLDVPINDSSNIGQYRLINCNSAVKQACSKVTPPIPIHYCAPAGFAILLKCRD
KEFNGTGECKNVSTVQCHGKIPVYSTOLLNGSLAESEVRIRSENLTNNAKTIIYQL
QLHNFRNKTIVFNHSGDLEVTTHSFNCGEFFCYNTSGLFNSTWANDNTSTQNDM
QSDNTITLPCRIKQIINWQVQAVYPIQGLRCDNSITGLLTRDGGEGNDTIG
KETRPTGDMNRNWRSELYKYKVEIKPLGVAPTAKRRVVEREKRAVGIGAVLLGF
LGAAGSTMGAASITLTVQAROLLSGIVQQNNLLKAIYVQOHLKLTWGIKQLQARV
LAVERYLKQQLLIGWCSGKLICTTVPWNVSNKTYNEIWENNMWIKWREID
NYQTITLIEESQNEKNEQDLALDKASLWNWFDISNLWYIRIFIMIVGGLIG
LRIIFAVLSIVNRCQGYSPLSQALTPSQGPDRPGGTKEGGEQGTSSIRLVSGF
LALAWDDLRLNCLRFILIVARIVETLGRGWESLKYLGKLNLAQYWGQELKSS
AISLLNATAIYAVAGTDRIIEVAHRAILNIPRIHQFFERALL"
8070..8693
/gene="nef"
8070..8693
/gene="nef"
/note="nef protein"
/codon_start=1
/product="nef"

```

```

/protein_id="AAF80538.1"
/db_xref="GI:8886639"
/translation="MGWNKSSIVGWPEIRERLRRTPTTAAAEVGAASKDLERHGA
ITSSNTSQNNAACAWLEAQDEGVFPVRPOVPLRPMTYKGAFDLSFFLKEGGDGL
IYSKKRQITLDLWVHTQCFDPDMQNTYTPGCIIRYPLCFGWCYKLVVPDPREVEANE
GENNCLLHPMSQHGIEDEDEVLRRKEDSSSLARRHIARELHPDFYKD"
BASE COUNT 3248 a 1564 c 2178 g 1945 t
ORIGIN
Query Match 5.9%; Score 19; DB 14; Length 8935;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 CTTTCATGTCCTTCTGGAG 129
DB 3572 CTTTCATGTCCTTCTGGAG 3554
Search completed: March 30, 2002, 07:17:31
Job time: 7243 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:19:22 ; Search time 1557.52 Seconds
(without alignments)
2207.772 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAAAAAATGTGAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	269	84.1	1842	12	AK017003	Mus muscu
2	107.4	33.6	544	10	AF247150	AF247150 AF247150
3	79.6	24.9	934	13	AQ744125	HS_5507_A
4	55.4	17.3	500	10	AW630609	h85a11.y
5	55.4	17.3	583	10	AW607690	MR3-HT048
6	48	15.0	442	11	BF835032	BF835032 RC3-HT086
7	47.2	14.8	545	10	BE693376	RC1-BT031
8	46.6	14.6	629	10	AW822729	AW822729 uq17g09.y
9	46	14.4	413	11	BF832095	PM3-HT090
10	46	14.4	966	11	BG167247	602342507
11	45.8	14.3	2903	12	AF077031	Homo sapi
12	44.4	13.9	452	10	AW820088	QVO-ST029

13	44.2	13.8	475	10	AW504119	UI-HF-BNO
14	43.8	13.7	296	10	BE763750	RC5-NT005
15	43.6	13.6	560	10	BE223530	kp81h08.y
16	43.6	13.6	564	10	BE580815	kp81h08.y
17	42.8	13.4	598	10	AV616209	AV616209
18	42.6	13.3	493	10	AI396994	fb08b12.y
19	42.2	13.2	475	10	AW839147	CM0-LT006
20	41.8	13.1	701	11	BG484180	602504794
21	41.6	13.0	355	11	BE830796	CM4-MT002
22	41.2	12.9	479	11	BF775055	285103 MA
23	40.2	12.6	300	10	AW890822	RC5-NT005
24	40.2	12.6	571	10	AW839966	MR4-LT007
25	40	12.5	649	11	BG829296	602753004
26	39.6	12.4	310	10	BE395495	601309832
27	39.2	12.3	495	10	AW503474	UI-HF-BNO
28	39.2	12.3	634	10	AW822725	uq17g02.y
29	38.8	12.1	584	10	AI155087	ud59a10.y
30	38.2	11.9	376	10	AA728886	nv37e08.f
31	38.2	11.9	444	13	AQ333088	HS_5010_A
32	38.2	11.9	679	13	AQ385814	RPC111-15
33	38.2	11.9	876	13	AZ674685	ENTF02TF
34	38.2	11.9	927	13	AZ675096	ENTF02TF
35	38	11.9	591	11	BG793262	UTSK SM10
36	38	11.9	622	10	AW822665	uq14D07.y
37	37.6	11.8	720	11	BF030754	601557471
38	37.4	11.7	796	10	AU132052	AU132052
39	37.4	11.7	811	10	AU132040	AU132040
40	37.2	11.6	545	10	AW877282	MR4-PT005
41	37.2	11.6	680	13	CNS02C2V	Tetraodon
42	37	11.6	411	11	BG726429	sad44f01.
43	37	11.6	711	10	AI827054	w107g04.x
44	37	11.6	926	11	BE799090	601588508
45	36.8	11.5	736	11	BE849002	uw01g10.y

ALIGNMENTS

RESULT: 1
AK017003 1842 bp mRNA HTc 05-JUL-2001
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933431F18, full insert sequence.
DEFINITION AK017003
ACCESSION AK017003.1
VERSION AK017003.1
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:4933431F18.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1842)
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 1842)
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 1842)
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Qy	61	TGTTGCAGTTACTGGCCGGTTTCTCTCAAGAGAGCCTTTTGGAAATTCAAACACTTTTCATGTC	120
Db	1092	TTTTACAGTTACTGGCCCAATTTCTCTCAAGAGAGCCTTTTGGAAATTCGAACACTTTTAGTGTCT	1151
Qy	121	CTTCTGGAGAACTTTTCAGATAAATCAGTATTTTGTCTATCCGATATTTCAAATTTGTGAAG	180
Db	1152	TTTCTGGAGACCTTTTCATGTAACCTCAATATTTTACCCTTCGAGTATTTTCAGATTTGTGAAG	1211
Qy	181	AAGTCCACAGAAAGAGTCTCTGTAAACACATTCGAGTTTCATCAATGGCCAGACCAT	240
Db	1212	AAGTCCACAGAAAGACCCATGTCTAAACACTTCGAGTTTCACCAAGTGGCCAGACCAT	1271
Qy	241	GGCACTCCT-GCCTCAGTAGATTTTTTCATCAAAATATGTCCTTATCTGAGGAGAGCCCA	299
Db	1272	GGCACTCCTGGCCCTCAGCAGATTTTTTTCATAAAATATGTCCTTATCTGAGGAGAGCCCA	1331
Qy	300	CATTACAGGACCCCTCCTTGT	320
Db	1332	CATTACAGGACCCCTCCTTGT	1352

RESULT 2

AF247150/c

LOCUS

DEFINITION

AF247150 544 bp mRNA EST 29-JAN-2001

AF247150 Oryctolagus cuniculus femoral head 2 months old

Oryctolagus cuniculus cDNA clone 61 similar to Mus musculus protein tyrosine phosphatase, mRNA sequence.

ACCESSION

AF247150

KEYWORDS

AF247150.1 GI:12597021

SOURCE

EST.

ORGANISM

Oryctolagus cuniculus

rabbits.

REFERENCE

1 (bases 1 to 544)

AUTHORS

Nikolaev,A.I., Dubovaya,V.I., Litvinov,D.Y., Poltarau,A.B., Prassolov,V.S., Ivanov,D.S., Amchenkova,A.M., Narovlyansky,A.N., Panasyuk,A.F. and Turpaev,K.T.

TITLE

Isolation of nitric oxide-induced genes from chondrocyte cDNA library using subtractive hybridization

JOURNAL

unpublished (2000)

COMMENT

Contact: Turpaev K

Engelhardt Institute of Molecular Biology

Vavilova 32, Moscow, 117984, Russia

Email: turpaev@genome.eimb.relarn.ru.

FEATURES

source

1. 544

/organism="Oryctolagus cuniculus"

/db_xref="taxon:9986"

/clone="61"

/clone_lib="Oryctolagus cuniculus femoral head 2 months old"

/tissue_type="femoral head"

/cell_type="primary articular chondrocyte"

/dev_stage="2 months old"

/note="breed: New Zealand White; obtained via subtractive hybridization of chondrocytes preincubated with 0.5 mM nitroglutathione vs. untreated cells"

BASE COUNT

173 a 102 c 103 g 166 t

ORIGIN

Query Match 33.6%; Score 107.4; DB 10; Length 544;

Best Local Similarity 73.6%; Pred. No. 4.8e-21;

Matches 167; Conservative 0; Mismatches 51; Indels 9; Gaps 2;

Qy 1 GAAATAATTGTAATGTTATTTGCTATCATACACAGAGATAGAAGTGGAGTTATCAAG 60

Db 359 GAAATAATTGTAATGTTATTTGCCATCATACACAGAGATAGAAGTGGCTTGTCAA 300

Qy 61 TGTTGCAGTTACTGGCCGGTTTCTCTCAAGAGAGCCTTTTGGAAATTCAAACACTTTTCATGTC 120

Db 299 TGCACCATTTACTGGCCAGTTCTTTCGAAGAGCCTTTGGAATTTGAACTTTTTC----- 243

```

Qy 121 CTTCTGGAGAACTTTCAGATTAAC-TCACTATTGTGTCATCCGAATATTTCAAAATGTGAA 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 ---CTGGAGAACTACCAGATACTTGAATCTTCATCATTTCAAAATGTTTAAATGTGAA 188
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 180 GAAGTCCACGAGAAAGAGTCACTCTGTAACACACTTGCAGTTCATCA 226
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 GAAGTCCATAAGTTTTTAAACTCAGCAGTAGATCTATGTCTCACCA 141

RESULT 3
LOCUS AQ744125 934 bp DNA GSS 16-JUL-1999
DEFINITION HS_5507_A2_B02_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-1083 Col-4 Row=C, DNA sequence.
ACCESSION AQ744125
VERSION   AQ744125.1 GI:5521647
KEYWORDS  GSS.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 934)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
           Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
           Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
           scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
           High Throughput Sequencing Center
           University of Washington
           401 Queen Anne Avenue North, Seattle, WA 98109, USA
           Tel: (206) 616-3618
           Fax: (206) 616-3887
           Email: jwallace@u.washington.edu
           Clones are derived from the human BAC library RPCI-11. For BAC
           library availability, please contact Pieter de Jong
           (pieter@dejong.med.buffalo.edu). Clones may be purchased from
           BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
           or from Resear h Genetics (info@resgen.com). BAC end web Server:
           http://www.htsc.washington.edu
           Plate: 1083 row: C column: 4
           Seq primer: T7
           Class: BAC ends
           High quality sequence stop: 934.
FEATURES
     source
       1..934
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="Plate=1083 Col-4 Row=C"
         /clone_lib="RPCI-11 Human Male BAC Library"
         /sex="male"
         /note="vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
           Male blood DNA was isolated from one randomly chosen donor
           and partially digested with a combination of EcoRI and
           EcoRI Methylase. Size selected DNA was cloned into the
           pBACe3.6 vector at EcoRI sites"
BASE COUNT 273 a 183 c 168 g 309 t 1 others
ORIGIN
Query Match 24.9%; Score 79.6; DB 13; Length 934;
Best Local Similarity 69.7%; Pred. No. 7.8e-13;
Matches 124; Conservative 0; Mismatches 49; Indels 5; Gaps 1;

Qy 1 GAAATAATTTGTAATGTTGCTATGATACACGAGAGATAGAGGTGGAGTTATCAAG 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 655 GGAATAATTTGTAATGTTGCTATGATACACGAGAGATAGAGGTGGAGTTATCAAA 714
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 TGTTCAGTACTGCGCGTTTCTCTGAAGGAGCGCTTTGGAATTCAAACATTTTCATGC 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 715 TGCTACCAATTACTGGCCCATTTCTCTGAAGAACCACTTGAATTTGAACACTTTCTGTGTTTC 774
Qy 121 CTTCTGGAGAACTTTCAGATAAAGTCACTTATTTGTCAATCCGAATATTTCAAAATGTGA 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 775 CT-----GGAGACTACCAGATACTTAATATTTTCATCATTTCAAAATGTTTCAAGTGTGA 827
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
LOCUS AW630609 500 bp mRNA EST 31-MAR-2000
DEFINITION hh85a11.y1 NCI-CGAP_GUI Homo sapiens cDNA clone IMAGE:2969564 5'
            similar to FR:Q15159 Q15159 TYROSINE PHOSPHATASE ;, mRNA sequence.
ACCESSION AW630609
VERSION   AW630609.1 GI:7377399
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
           National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
           Tumor Gene Index
           Unpublished (1997).
JOURNAL   Other_ESTs: hh85a11.x1
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
           Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
           Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
           I.M.A.G.E. Consortium DNA Sequencing by: Washington University
           Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           image.llnl.gov/image/html/iresources.shtml
           Seq primer: -40RP from Gibco
           High quality sequence stop: 425.
FEATURES
     source
       1..500
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:2969564"
         /clone_lib="NCI-CGAP_GUI"
         /tissue_type="2 pooled high-grade transitional cell
           tumors"
         /lab_host="DH10B"
         /note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
           Site_1: SalI; Site_2: NotI; Cloned unidirectionally.
           Primer: Oligo_dT. Library constructed by Life
           Technologies."
BASE COUNT 149 a 116 c 118 g 117 t
ORIGIN
Query Match 17.3%; Score 55.4; DB 10; Length 500;
Best Local Similarity 50.5%; Pred. No. 8.3e-06;
Matches 162; Conservative 0; Mismatches 156; Indels 3; Gaps 1;

Qy 1 GAAATAATTTGTAATGTTGCTATGATACACGAGAGATAGAGGTGGAGTTATCAAG 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 141 GAGCAAAATCCACAGTGTAGCCATGATCAAGAGTAGAGGAGAGAAAATCAAA 200
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 TGTTCAGTACTGCGCGTTTCTCTCT---GAAGGAGCGCTTTGGAATTCAAACACTTTTCAT 117
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 201 TGCCAGCGCTATTGGCCCAACATCTCTAGGCAAAACAACAATGGTCAGCAACAGACTTGA 260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 118 GTCCTTCGGAGAACTTTCAGATAACTCAGTATTTTGTTCATCCGGAATATTTCAAATTTGG 177
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 CTGGCTCTTGTGAAATCGACAGCTGAAGGCGCTTTGTGGTGAGGCAATGACCCCTTGA 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 178 AAGAAGTCCACAGAAAGAGTCACTCTCTGTAACACACTTTCAGTTTCATCAATGGCCAGAC 237
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 GATATTCAGACACAGAGGTGGCCCATATTTCTCTCATCTGAATTTTCATCGCTGGCCGAC 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

QY 76 CCCGTTCTCTCAAGGAGCCTTTGGAATTCACACACTTTTCATGTCTCTCTGGAGAACATTT 135
Db 112 CTTTGAAGAAAGACTCTCGGATCCGATTTGGCTTCCTCACAGTGACCCCATCTAGGCGTG 171
QY 136 CAGATAACTCAGTATTTTGTCTATCCGAATATTTCAAAATTTGTAAGAAGTCCACAGGAAAG 195
Db 172 GAGACATGATCATATATAAGAAACAGCGCTAGAAATTCACACAGAGAACGGCAG 231
QY 196 AGTCACTCTGTAACACACTGTCAGCTTCATCAATCAAGCCAGACCATGGGCACTCTGCCCTCA 255
Db 232 AAACGCCAGGTGACCCACTTCCAGTCTTGTAGCTGCCAGACATAGTGTCTCCTCTCTCA 291
QY 256 GTAGATTTTTCATCAATATATGTCGTTATGTGAGAGAGAGAGCCACATTT 303
Db 292 GCAGCTTCCCTCATTTGACTTCTTGAGAGTGGTTCAGAAACACGACAGAT 339

RESULT 7
LOCUS BE693376 545 bp mRNA EST 11-SEP-2000
DEFINITION RC1-BT0314-210700-017-b09 BT0314 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE693376
VERSION BE693376.1 GI:10080536
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 545)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=st2-RC1-BT0314-210
700-017-b09&t3=2000-07-21&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 66
High quality sequence stop: 543.
FEATURES
Location/Qualifiers
source
1..545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="BT0314"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 155 a 131 c 126 g 133 t
ORIGIN

Query Match 14.8%; Score 47.2; DB 10; Length 545;
Best Local Similarity 51.1%; Pred. No. 0.0022;

```

```

Matches 140; Conservative 0; Mismatches 128; Indels 6; Gaps 1;
QY 1 GAAATAATTTGTAATTTGTTATTTGCTATGATAACACAGAGATAGAAGTGGAGTTATCAAG 60
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 141 GAGAAAATGTATATGCCATCATTTATGTTGACTAAATGTTGAACAGGGAAGAACCAA 200
Db 141 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 161 TGTTCGAGTTACTGGCCGCTTTCTCTGAAGGAGCCCTTTGGAAATTCACACACTTTTCATGTC 120
Db 161 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 201 TGTGAGGAGTATTTGGCCC-----TCCAAGCAGGCTCAGGACTATGGAGACATAACTGTG 254
Db 201 111 111 111 111 111 111 111 111 111 111 111 111
QY 121 CTCTGGAGAACTTTTCAGATAACTCAGTATTTTGTCTCCGAAATATTTTCAAAATTTGGAAG 180
Db 121 111 111 111 111 111 111 111 111 111 111 111
QY 255 GCAATGACATCAGAAATTTGTTCTTCGGAAATGGACCATCAGAGATTTACAGTGAANAAT 314
Db 255 111 111 111 111 111 111 111 111 111 111 111
QY 181 AAGTCCACAGGAAAGAGTCACTCTGTAAACACATTTGCAGTTTCATCAAAATGGCCAGACCAT 240
Db 181 111 111 111 111 111 111 111 111 111 111 111
QY 315 ATCCAGACAAAGTGAGAGTCACCCCTCTGAGACAGATTTCCATTTCCACCTCCTGGCCAGACCAC 374
Db 315 111 111 111 111 111 111 111 111 111 111 111
QY 241 GGCACCTCTCGCTCAGTAGATTTTTCATCAAAAT 274
Db 241 111 111 111 111 111 111 111 111 111 111
QY 375 GGTGTTCCCGACACCACTGACCTGCTCATCAACT 408
Db 375 111 111 111 111 111 111 111 111 111 111

RESULT 8
LOCUS AW822729 629 bp mRNA EST 17-MAY-2000
DEFINITION uq17g09.y1 Ren Stubbs mouse thymus Mus musculus cDNA clone
IMAGE:2802784 5' similar to gb:X63440 M.musculus mRNA for
p19-protein tyrosine phosphatase (MOUSE);, mRNA sequence.
ACCESSION AW822729
VERSION AW822729.1 GI:7915806
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 629)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LILN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1042308
Seq primer: Primer name ambiguous
High quality sequence stop: 464.
Location/Qualifiers
source
1..629
/organism="Mus musculus"
/db_xref="C3H"
/db_xref="taxon:10090"
/clone.lib="IMAGE:2802784"
/clone.lib="Ren Stubbs mouse thymus"
/sex="mixed"
/dev_stage="3 weeks"
/lab_host="DH10B"
/note="Organ: thymus; Vector: pT73D-Pac; Site_1: NotI;
Site_2: PacI; 1st strand cDNA was primed with an oligo(dT)
primer; double-stranded cDNA was ligated using 5' linker
ggccgctat and 3' linker aactggagcttaatt. Library is
size-selected >2.5 kb and average insert size is 3.5 kb.
Clones were arrayed from primary plating; non-amplified.
Library constructed by X. Ren and L. Stubbs (Lawrence
Livermore National Laboratory and DOE Joint Genome

```

BASE COUNT 183 a 118 c 152 g 175 t 1 others
ORIGIN Institute, 7000 East Ave, L-453, Livermore, CA 94550).

Query Match 14.68; Score 46.6; DB 10; Length 629;
Best Local Similarity 49.88; Pred. No. 0.0034;
Matches 149; Conservative 0; Mismatches 144; Indels 6; Gaps 1;

QY 1 GAAATAATTTGTAATGTTATGCTATGATACACAGAGAGATAGAGGTGGAGTTATCAAG 60
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 247 GAGTATATGTTGATCATGCTGCTGCGAGATTTGAGATGGGAGGAAAG 306
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 61 TGTTCAGTACTGGCCGCTTCTCTGAAGAGGCTTTGGAATTCACACATTCATGTC 120
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 307 TGTGAGGCTACTGGCTTGTATGGAGAGATCTATAACATTTGCACCATTTAAAT 366
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 121 CTCTGAGACTTTCAGATACTCAGTATTTGTCATCCGATATTTCAATTTGTCAG 180
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 367 TCTTGTGAATGAACAGACGACGACTCTTCATCCGACACTTTTACTTTGAATTT 426
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 181 AAGTCCACAGAAAGACTCTGTAAAACACTTGCAGTTTCATCAATGCCAGACCAT 240
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 427 CAAAATGAATCCCGCTGCTCT-----ATCAGTTTCATTACGTAAGTGGCCAGACCAT 480
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 241 GGCACCTGCTCAGTAGATTTTTCATCAATATGTCGCTTATGTGAGAGAGCCA 299
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 481 GATGTCCTCGTCATTTGATTTCTTCTGGACATGATAAGCTTAATGAGGAATAACCA 539
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

RESULT 9
BF832095/c
LOCUS BF832095 413 bp mRNA EST 13-JAN-2001
DEFINITION PM3-HT0909-181000-014-f02 HT0909 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF832095
VERSION BF832095.1 GI:12180468
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 413)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-HT0909-
181000-014-f02&t3=2000-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 2
High quality sequence stop: 413.
Location/Qualifiers
1. .413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0909"
/dev_stage="Adult"

FEATURES
source
1. .966
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4452672"
/clone_lib="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.

/note="Organ: head,neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 105 a 96 c 116 t
ORIGIN

Query Match 14.48; Score 46; DB 11; Length 413;
Best Local Similarity 50.08; Pred. No. 0.0045;
Matches 143; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

QY 36 AGAGATAGAGGTGGAGTTATCAAGTGTTCAGTTACTGGCCCTTTCTCT---GAAGGA 92
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 413 AGAGTAGAGGAGAAATCAATGCCAGGCTATTTGCCCAACATCTTAGGCAAAAC 354
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 93 GCCTTTGGAATTCACACATTTTCATGTCCTTCTGGAGAACCTTCAGATACTAGTATT 152
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 353 AACAAATGCTCAGCAACAGACTTCGACTGGCTCTTGTGAGATGCGAGCTGAAGGGCTT 294
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 153 TGTTCATCGGATATTTCAAAATTTGGAAGAGTCCACAGGAAGAGTCACTCTGTAAACA 212
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 293 TGTGGTGAAGGCAATGACCTTTGAAGATATTCAGACAGAGAGGTGGCCCATATTTCTCA 234
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 213 CTTCAGCTTCATCAATAGGCGACACCATGGCCTCTGCTCAGTACAGTATTTTTCATCAA 272
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 233 TCTGAATTTCACTGCTGGCCAGACCATGATACACCTTCTCAACCATGATGCTGTAC 174
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
QY 273 ATATGTCGGTATGTGAGGAGAGCCACATTCACAGGACCCCTCTT 318
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
Db 173 TTTTATCTCTACATGATGAGACATCCACAGATCAGGCCCAATCCAT 128
II IIIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

RESULT 10
BG167247
LOCUS BG167247 966 bp mRNA EST 06-FEB-2001
DEFINITION 602342507F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452672 5',
mRNA sequence.
ACCESSION BG167247
VERSION BG167247.1 GI:12673950
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 966)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10241 row: n column: 01
High quality sequence stop: 714.
Location/Qualifiers
1. .966
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4452672"
/clone_lib="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.

Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

283 a	226 c	232 g	225 t
BASE COUNT			

[illegible]

RESULT	11
AF077031	
LOCUS	2903 bp mRNA HTC
DEFINITION	Homo sapiens protein tyrosine phosphatase homolog mRNA, complete cds.
ACCESSION	AF077031
VERSION	AF077031.1 GI:4689109
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2903)
AUTHORS	Liu,T., Zhang,J., Fu,G., Zhang,Q., Ye,M., Zhou,J., Wu,J., Shen,Y., Yu,M., Chen,S., Mao,M. and Chen,Z.
TITLE	Human protein tyrosine phosphatase (70zpep) homolog Unpublished
REFERENCE	2 (bases 1 to 2903)
AUTHORS	Liu,T.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-1998) Shanghai Institute of Hematology, Shanghai Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II, Shanghai 200025, P.R. China
FEATURES	Location/Qualifiers

```

1. .2903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NB4"
73. 2496
/codon_start=1
/product="protein tyrosine phosphatase homolog"
/protein_id="AB27764.1"
/db_xref="GI:4689110"
/translation="MDQRIQLQFLDEAQSCKITKEEFANEFLLKROSTKYKADKTY
PTVAEPKPNIKKNRYKDILPYDYSVEISLITSDSSYINANFKGVYGPAYLAT
QGPLSTDYIWRNINWEYSVLLIGMACMEYEMGKKCRKEWAEPMQLEFGPFSVC
EAERKSDYIIRTLKVFNSERTTYIOFYHKNPWHDVPSIDPILLEIWDVRCYOEID

```

BASE COUNT
ORIGIN

955 a 580 c 528 g 840 t

DSVPTCIHCSAGGRTGVCAIDYTWMLKDGIIIPENFSVFSLLIEMRTQRPSSLVOTQ
EOYELYNVAVLELFRQMDVIRKHSGETSQAKHCIPKKNHTLOADSYPNLPKSTTT
AAKMNQOQRTKWEIKESSDFRTEISAKAEELVHPAKSSTFDFLEUNSEFDKNAD
PTMKWQTKAFLPIVGELOKHQSLDGLSPFEGCNSKVPYNAAGRVNSKVPITRKST
PRELIOQRTKEVDKSEKNSYI LESQPHDSCFVEMQQAQKVHVSSAELSLBPTDSKHQ
IRNASNKHSDSALGYSYI ELVENYFESSPPGTSKMSLDLPEKODGVTFPSSL
LPTSSYFYNHSDLSNPTINISLLNGEASVATAPRIDDEIPPLPVMWTPES
FIVEPAGFSPNPKLSASVKVYIGTSLGEGTSEPKKFDSDVILRPSKSVKLRRSP
KSELHODRSSPPPLPRTLLEFFLADECMQAQSIETYSTYSPOTMENSTSSKOTLK
TSGKSTSRSKSLKILRNMKKSCICNCSPPNKPAAESVQSSNNSSSFLNFGFANRFSKPGP
RNPPTWNI*

Query Match	14.3%	Score 45.8	DB 12	Length 2903
Best Local Similarity	50.2%	Pred. No. 0.0085		
Matches 144	Conservative 0	Mismatches 137	Indels 6	Gaps 1
Qy	1	GAATAAATTTGTAATGTTATTGTCTATGATACACGAGAGATAGAAGGTGGAGTTCATCAAG	60	
Db	427	GAATATAGTCCTTATCATTTGGTATGGCATGCATGGAGTATGAAATGGGAAGAAAAG	486	
Qy	61	TGTTGCAGTTACTGGCCCGTTTCTCTGAAGAGCGCTTTGGAATTCACAACTTTTCATGTC	120	
Db	487	TGTGAGCGCTACTGGGCTGAGCGAGTAGAGATGCAGCTGGAAATTTGGCCCTTTCTCTGTA	546	
Qy	121	CTTCTGGAGCAACTTTCAGATAACTCAGATATTTTGTCTCCGAATATTTTCAAATTTGTGAAG	180	
Db	547	TCCTGTGAAGCTGAAAAAAGGAATCTGATTATATATCAGGACTCTAAAAGTT-----	600	
Qy	181	AGTCCACAGGAAGAGTCACTCTGTAAAACACTTGCAGTTTCATCAAATGGCCAGACCAT	240	
Db	601	ANGTTCAATAGTGAACCTGCAACTATCTCCAGTTTCATTACAGAATTTGGCCAGACCAT	660	
Qy	241	GGCACTCCCTGCCTCAGTAGATTTTTTTCATCAAAATATGTCGGTTATGT	287	
Db	661	GATGTACCTTCATCTATAGACCCTATTCTTGGAGTCATCTCGGATGT	707	

RESULT	: 12
AW820088/c	
LOCUS	
DEFINITION	QVO-ST0294-070400-186-e1l ST0294 Homo sapiens cDNA, mRNA sequence.
ACCESSION	AW820088
VERSION	AW820088.1 GI:7913082
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria;	Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 452)	
Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.	
REFERENCE	
AUTHORS	

JOURNAL
MEDLINE
COMMENT

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A. J. G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/getbtlm2.pl?pl1=512-0V0-570294-070>)

400-186-ell1t3-2000-04-07t4-1)
Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 452.
Location/Qualifiers

FEATURES

source

1. .452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0294"
/dev_stage="Adult"

/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 96 a 121 c 120 g 115 t
ORIGIN

Query Match 13.9%; Score 44.4; DB 10; Length 452;

Best Local Similarity 47.5%; Pred. No. 0.014;

Matches 132; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 26 TGATACACAGAGATAGAGGTGGAGTTATCAAGTGTTCAGTGTACTGGCCGTTTCTC 85

Db 452 TGACACCCGCTTTGAGGAAGGCGGAGGAAAGTGTGGCCAGTACTGGCCCTTTAGAAA 393

QY 86 TGAAGGAGCCTTTGGAAATCAACACACTTTCATCTCTCTGGAGAACTTTTCAGATACTC 145

Db 392 AGACTCTCGATCCGATTTGGCTTCTCAGCTGACCAATCTAGCGGTGGAGAACATGA 333

QY 146 AGTATTTGTATCCGAATATTTCAAAATTTGAAGAGTCCACAGAGAAAGAGTCACTCTG 205

Db 332 ATCATTAAGAAACACACCTAGAAATTCACACACAGAGGAGCGCAGAACGCCAGG 273

QY 206 TAAACACTTGCAGTTCATCAATAGGCGACACCATGCGCTCGCTCAGTAGATTTT 265

Db 272 TGACCCACTTCCAGTCTTCCAGGTGGCAGACTATGTGTCCTCTCCTCAGCAGCTTCCC 213

QY 266 TCATCAATATATGTCGCTTATGTGAGGAAGCCACATT 303

Db 212 TCATTGACTTCTTGAGAGTGGTCAGAAACCCAGCAGT 175

RESULT 13

AW504119

LOCUS

AW504119 475 bp mRNA EST 02-MAR-2000

UI-HF-BN0-ali-h-10-0-UI.rl NIH_MGC_50 Homo sapiens cDNA clone

IMAGE:3079915 5', mRNA sequence.

AW504119.1 GI:7141786

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 475)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabps-remail.nih.gov

Eco RI site shown at the beginning of the sequence.

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library preparation: M.B. Soares Lab

cDNA Library Arrayed by: M.B. Soares Lab

DNA Sequencing by: M.B. Soares Lab

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward.

FEATURES

source

1. .475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3079915"
/clone_lib="NIH_MGC_50"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"

/note="Vector: p7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4Kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 134 a 87 c 115 g 139 t

ORIGIN

Query Match 13.8%; Score 44.2; DB 10; Length 475;

Best Local Similarity 49.8%; Pred. No. 0.016;

Matches 143; Conservative 0; Mismatches 138; Indels 6; Gaps 1;

QY 1 GAAATAAATTGTAATCTTATTGCTATCATACACAGAGATAGAGTGGAGTTCATCAAG 60

Db 76 GAATATAGTCTCTTATCATTTGTTATGGCATGATGAAATGGGAAAGAAAAG 135

QY 61 TGTTCGAGTTACTGGCCGCTTCTCTGAAGGAGGCTTTGGAATTCAAACACACTTTTCATGTC 120

Db 136 TGTGCGGCTACTGGCTGAGCCAGGAGAGATGCGCTGGAATTTGCCCTTTCTCTGTA 195

QY 121 CTCTCGAGAACTTTCAGATACTCAGTATTTTGTCTATCCGAATATTTCAAAATTTGCGAAG 180

Db 196 TCCTGTGAAGCTGAAAAAGGAAATCTGATTATATATACAGACTCTAAAAAGTT----- 249

QY 181 AACTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAATGCCAGACCAT 240

Db 250 AAGTTCAATAGTGAAGTCTGAACTCTACCACTTCTACCAAGTTCATTACAGAAATGCCAGACCAT 309

QY 241 GGCACCTCTCCCTCAGTAGATTTTTCATCAAAATATGTCGGTTATGT 287

Db 310 GATGACCTTCATCTATAGACCCCTATTCTTGAGCTCATCTGGGATGT 356

RESULT 14

BE763750

LOCUS

BE763750 296 bp mRNA EST 19-SEP-2000

RC5-NT0053-140600-022-D02 NT0053 Homo sapiens cDNA, mRNA sequence.

BE763750

BE763750.1 GI:10193674

ACCESSION

VERSION

KEYWORDS

SOURCE

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 296)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,

Goldman, G.H., Carvalho, A.F., Matsuoka, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org, br
This sequence was derived from the FAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC5-NT0053-140
600-022-0026t3-2000-06-14&t4-1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 296.

FEATURES

source

Location/Qualifiers
1..296
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0053"
/dev_stage="adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
78 a 82 c 68 g 68 t

BASE COUNT

ORIGIN
218 a 63 c 98 g 181 t

Query Match 13.7%; Score 43.8; DB 10; Length 296;
Best Local Similarity 43.4%; Pred. No. 0.018;
Matches 114; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
Qy 88 AAGGAGCCTTTGGAAATTCACACATCTTCCTCTGGAGAACTTCAGATAACTCAG 147
Db 32 AAACGACAATGGTCAGACAGACCTTCGACTGCTCTGTGAGNATCCAGCCTGAAG 91
Qy 148 TATTTTGCATCCGAATATTTCAATTTGTAAGTGTGAAGTCCACGAGGAAGTCACCTCTGTA 207
Db 92 GCGTTTGGTGGGCAATGACCTTGAAGATATTCAGCCAGAGAGGTGGCCATATT 151
Qy 208 AAACATTCGAGTTCATCAATGGCCAGCATGCCACTCTCGCTCAGTAGATTTTTC 267
Db 152 TCTCATCTGAATTTACCTGCGCTGGCCAGACCATGATACACCTTCTCAACAGATGATCTG 211
Qy 268 ATCAAAATGTCGCGTTATGTGGAAGAGCCACATTCAGGAGCCCTCCT 318
Db 212 CTACTTTTATCTCTACATGAGACATCCAGACATCAGGCCCCATATT 262

RESULT 15

BE223530 560 bp mRNA EST 09-MAY-2001
LOCUS kp81h08.v1.TB95TM-SSFH Strongyloides stercoralis cDNA 5' similar
DEFINITION to TR:Q15718 Q15718 PTPSIGMA PRECURSOR ;, mRNA sequence.
ACCESSION BE223530
VERSION BE223530.1 GI:8928766
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloidea; Strongyloidea.
1 (bases 1 to 560)

REFERENCE

AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, R.,
Ronko, J., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
Wilson, R.

TITLE

The Washington Univ. Nematode EST Project, 1999

JOURNAL

Unpublished (1999)

COMMENT

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810
Email: estewatson.wustl.edu

The library was constructed by Dr. Thomas Nutman and colleagues of
NIAID, NIH (nutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.

Seq primer: -40RP from Gibco

High quality sequence stop: 424.

FEATURES

source

Location/Qualifiers
1..560
/organism="Strongyloides stercoralis"
/strain="Fillariform larvae obtained from humans"
/db_xref="taxon:6248"
/clone_lib="TB95TM-SSFH"
/lab_host="XL-1 Blue MRF" (Stratagene)"
/note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 4 x 10B5
fillariform larvae which had been isolated from infected
humans. cDNA was constructed and, using adaptors, was
cloned unidirectionally into the vector from the EcoRI
site to the XhoI site. The library has an amplified
titer of 1.5 x 10E6 pfu/ml and an amplified, undiluted
titer of 7 x 10E9 pfu/ml. The average insert size of the
unamplified library is 975 bp (range, 500-1500)."
218 a 63 c 98 g 181 t

BASE COUNT

ORIGIN
218 a 63 c 98 g 181 t

Query Match 13.6%; Score 43.6; DB 10; Length 560;
Best Local Similarity 50.0%; Pred. No. 0.025;
Matches 140; Conservative 0; Mismatches 134; Indels 6; Gaps 1;
Qy 1 GAAATATATTGTAATGTTATGCTATGATACCCAGAGATAGAGGTGGAGTTATCAAG 60
Db 54 GAAACAGAAATCATCATGTTATAGTTATGTTAAACAAAATTAAGAGTAGAAAAA 113
Qy 61 TGTTCAGTTACTGGCCCGCTTCTCTGGAAGAGCCCTTTGGAAATTCAAACACATTCATGTC 120
Db 114 AGTTATGAATTTGGCCATCAGAAAAAGTTATCAATATGGAATTTTGTGTAACCT 173
Qy 121 CTTCTGGAGAACTTTTCAGATAACTCAGTATTTTGTTCATCCGGAATATTTCAAAATTTGGAAG 180
Db 174 ATTGCTGA-----ATATAATATGCAATATATATATTAAGAGAAATTTAAATGAATGAT 227
Qy 191 AAGTCCAGGAAAGTCACCTCTGTAACACATTCGAGTTCATCAATGGCCAGACCAT 240
Db 228 ACACAATCAGGAAATTTATAAAACAATAAGACATTTTCAATATATGATTTGGCCAGAACAT 287
Qy 241 GGCACCTCTGCTCAGTAGATTTTTCATCAATATATGTC 280
Db 238 GGTGTACCAGAGTCAGCTGAATGTTTGAATTTGTAC 327

Search completed: March 30, 2002, 04:50:05

Job time: 1843 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:21:32 ; Search time 88.7 Seconds
(without alignments)
817.057 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAAATAATGTAATGCTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	17.8	8040	1	US-08-596-291-1
2	57	17.8	8040	3	PCT-US94-09943-1
3	57	17.8	8043	5	PCT-US94-09943-1
4	57	17.8	8119	4	US-09-230-640-45
5	47.2	14.8	5117	3	US-08-854-585-1
6	47.2	14.8	5117	5	PCT-US95-05512-1
7	44.2	13.8	3580	4	US-09-081-345-1
8	38.8	12.1	285	1	US-08-202-389-3
9	36.6	11.4	1210	4	US-08-811-481-30
10	36.6	11.4	1413	2	US-08-239-276-9
11	36.6	11.4	1413	3	US-08-468-579B-9
12	36.6	11.4	1413	2	US-08-468-577B-9
13	36.6	11.4	3311	2	US-08-239-276-10
14	36.6	11.4	3311	2	US-08-468-579B-10
15	36.6	11.4	3311	3	US-08-468-577B-10
16	34.4	10.8	2351	2	US-08-548-159-6
17	33.8	10.6	3229	1	US-07-777-715-8
18	33.8	10.6	3229	3	US-08-170-126-3
19	33.8	10.6	3229	3	US-08-954-418-3
20	33.2	10.4	5690	2	US-08-447-464-2
21	33.2	10.4	5690	2	US-08-716-679-2
22	33	10.3	33	4	US-09-081-345-9
23	33	10.3	33	4	US-09-081-345-10
24	32.8	10.2	3969	1	US-08-241-853-1
25	32.8	10.2	3969	2	US-08-850-917-1
26	32	10.0	432	2	US-09-001-944-11
27	32	10.0	432	4	US-09-240-004A-11

c 28	32	10.0	435	2	US-09-001-944-9	Sequence 9, Appli
c 29	32	10.0	435	4	US-09-240-004A-9	Sequence 9, Appli
c 30	32	10.0	450	4	US-09-240-004A-13	Sequence 13, Appl
c 31	32	10.0	498	2	US-09-001-944-1	Sequence 1, Appli
c 32	32	10.0	498	2	US-09-001-944-3	Sequence 3, Appli
c 33	32	10.0	498	2	US-09-001-944-5	Sequence 5, Appli
c 34	32	10.0	498	2	US-09-001-944-7	Sequence 7, Appli
c 35	32	10.0	498	4	US-09-240-004A-1	Sequence 1, Appli
c 36	32	10.0	498	4	US-09-240-004A-3	Sequence 3, Appli
c 37	32	10.0	498	4	US-09-240-004A-5	Sequence 5, Appli
c 38	32	10.0	498	4	US-09-240-004A-7	Sequence 7, Appli
c 39	30.8	9.6	270	2	US-08-446-345-25	Sequence 25, Appl
c 40	30.8	9.6	270	2	US-08-446-345-31	Sequence 31, Appl
c 41	30.8	9.6	5162	2	US-08-916-917-13	Sequence 13, Appl
c 42	30.8	9.6	5162	3	US-09-225-170-13	Sequence 13, Appl
c 43	30.4	9.5	1370	4	US-09-026-408-12	Sequence 12, Appl
c 44	30.4	9.5	1371	4	US-09-026-408-1	Sequence 1, Appli
c 45	30.4	9.5	2852	3	US-09-027-137-2	Sequence 2, Appli

ALIGNMENTS

RESULT : 1
US-08-536-291-1
Sequence 1, Application US/08596291
Patent No. 5821075
GENERAL INFORMATION:
APPLICANT: GONZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,291
FILING DATE: 09-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LC461/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8040 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: HOMO SAPIENS
NAME/KEY: CDS
LOCATION: 78..7475
US-08-596-291-1

Query Match 17.8%; Score 57; DB 1; Length 8040;
Best Local Similarity 50.8%; Pred. No. 1.8e-08;
Matches 163; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

QY 1 GAAATAATTTGTAATGTTATGCTATGATACCAAGAGAGATAGAGGTGGAGTTTCAAG 60
DB 6912 GAGCAAAAATCCACAGTAGCATGACTCAAGAAGTAGAAGAGAGAAAAATCAAA 6971
QY 61 TGTGTCAGTTACTGGCCGTTTCTCT---GAAGGAGCGTTTGGAAATTCAAACACTTTTCAT 117
DB 6972 TGCAGCGCTATTGGCCCAACATCTAGGCAAAACAACAATGTCAGCAACAGACTTCGA 7031
QY 118 GTCCCTTCGGAGAACTTTCAGATACTCAGTATTTTGTATCCGGAATATTTCAAAATTTGTG 177
DB 7032 CTGGCTCTTGTGAGAAATGCAAGAGCTGAAGGGCTTTTGTGAGGGCAATGACCCCTTGA 7091
QY 178 AAGAGTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAATGGCCAGAC 237
DB 7092 GATATTTCAGACAGAGAGGTGGCCCATATTTCTCATCAAAATATGTCGGTTATGTCAGTGCCTGGCCAGAC 7151
QY 238 CATGGCACTCCTGCTCAGTAGATTTTTCATCAAAATATGTCGGTTATGTCAGTGCCTGGCCAGAC 297
DB 7152 CATGATACACTTCTCAACAGATGATCTGCTTACTTTTATCTCCTACATGAGACATC 7211
QY 298 CACATTACAGACCCCTTCCTT 318
DB 7212 CACAGATCAGGCCCAATCATT 7232

RESULT 2

US-09-100-804-1
Sequence 1, Application US/09100804
Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESSON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK

TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LO461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8040 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
FEATURE:
NAME/KEY: CDS
LOCATION: 78..7475
US-09-100-804-1

Query Match 17.8%; Score 57; DB 3; Length 8040;
Best Local Similarity 50.8%; Pred. No. 1.8e-08;
Matches 163; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

QY 1 GAAATAATTTGTAATGTTATGCTATGATACCAAGAGAGATAGAGGTGGAGTTTCAAG 60
DB 6912 GAGCAAAAATCCACAGTAGCATGACTCAAGAAGTAGAAGAGAGAAAAATCAAA 6971
QY 61 TGTGTCAGTTACTGGCCGTTTCTCT---GAAGGAGCGTTTGGAAATTCAAACACTTTTCAT 117
DB 6972 TGCAGCGCTATTGGCCCAACATCTAGGCAAAACAACAATGTCAGCAACAGACTTCGA 7031
QY 118 GTCCCTTCGGAGAACTTTCAGATACTCAGTATTTTGTATCCGGAATATTTCAAAATTTGTG 177
DB 7032 CTGGCTCTTGTGAGAAATGCAAGAGCTGAAGGGCTTTTGTGAGGGCAATGACCCCTTGA 7091
QY 178 AAGAGTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAATGGCCAGAC 237
DB 7092 GATATTTCAGACAGAGAGGTGGCCCATATTTCTCATCAAAATATGTCGGTTATGTCAGTGCCTGGCCAGAC 7151
QY 238 CATGGCACTCCTGCTCAGTAGATTTTTCATCAAAATATGTCGGTTATGTCAGTGCCTGGCCAGAC 297
DB 7152 CATGATACACTTCTCAACAGATGATCTGCTTACTTTTATCTCCTACATGAGACATC 7211
QY 298 CACATTACAGACCCCTTCCTT 318
DB 7212 CACAGATCAGGCCCAATCATT 7232

RESULT 3

PCT-US94-09943-1

Sequence 1, Application PC/TUS9409943

GENERAL INFORMATION:

APPLICANT:
APPLICANT:
APPLICANT:
APPLICANT:

TITLE OF INVENTION:

TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: TWOMEY, MICHAEL J.
REGISTRATION NUMBER: P-38,349
REFERENCE/DOCKET NUMBER: L0461/7000W0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8043 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: HOMO SAPIENS
FEATURE:
NAME/KEY: CDS
LOCATION: 78..7478
PCT-US94-09943-1

Query Match 17.8%; Score 57; DB 5; Length 8043;
Best Local Similarity 50.8%; Pred. No. 1.8e-08;
Matches 163; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

QY 1 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAGGTGGAGTTATCAAG 60
DB 6915 GAGCAAAATCCACAGTATGATGCTGATGCTCAAGAGTAGAGGAGAAATCAAA 6974
QY 61 TGTTCAGTTACTGGCCCTTCTCTCT---GAAGGAGCCTTTGGAATTCAAACACTTTCAT 117
DB 6975 TGCCAGCGCTATTGGCCCAACATCCTTAGGCAAAACAAATGTCAGCAACAGACTTCA 7034
QY 118 GTCCCTTCTGAGAACTTTCAGATACTCAGTATTTGTCATCGAATATTTCAAAATTGTG 177
DB 7035 CTGGCTCTTGTGAGAAATGACAGCTGAAGGGCTTTGTTGGTGGGGCAATGACCCCTTGA 7094
QY 178 AAGAGTCCACAGGAAAGAGTCACTCTGTAAACACTTTCAGTTTCATCAAAATGGCCAGAC 237
DB 7095 GATATTCAGACAGGAGGTGGCCATATTTCTCATCTGAATTTCACTGCTGGCCAGAC 7154
QY 238 CATGGCACTTCCTCCTCAGTATGTTTTTTCATCAAAATATGTCGTTATGTGAGGAGAGC 297
DB 7155 CATGATACACCTTCTCAACAGCATGATCTGCTTACTTTTCTCTCTACATGATGACACATC 7214
QY 298 CACATTACAGGACCCCTCCTT 318
DB 7215 CACAGATCAGGCCCAATCAAT 7235

RESULT 4
US-09-290-640-45
Sequence 45, Application US/09290640
Patent No. 6204055
GENERAL INFORMATION:
APPLICANT: Dean, Nicholas M.
APPLICANT: Marcusson, Eric G.
TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling

FILE REFERENCE: ISPH-0351
CURRENT APPLICATION NUMBER: US/09/290,640
CURRENT FILING DATE: 1999-04-12
NUMBER OF SEQ ID NOS: 85
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 45;
LENGTH: 8119;
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (64)..(7521)
PUBLICATION INFORMATION:
JOURNAL: FEBS Lett.
VOLUME: 337
ISSUE: 2
PAGES: 200-206
DATE: 1994-01-10
DATABASE ACCESSION NUMBER: D21209/Genbank
DATABASE ENTRY DATE: 1999-02-05
US-09-290-640-45

Query Match 17.8%; Score 57; DB 4; Length 8119;
Best Local Similarity 50.8%; Pred. No. 1.8e-08;
Matches 163; Conservative 0; Mismatches 155; Indels 3; Gaps 1;
QY 1 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAGGTGGAGTTATCAAG 60
DB 6958 GAGCAAAATCCACAGTATGATGCTGATGCTCAAGAGTAGAGGAGAAATCAAA 7017
QY 61 TGTTCAGTTACTGGCCCTTCTCTCT---GAAGGAGCCTTTGGAATTCAAACACTTTCAT 117
DB 7018 TGCCAGCGCTATTGGCCCAACATCCTTAGGCAAAACAAATGTCAGCAACAGACTTCA 7077
QY 118 GTCCCTTCTGAGAACTTTCAGATACTCAGTATTTGTCATCGAATATTTCAAAATTGTG 177
DB 7078 CTGGCTCTTGTGAGAAATGACAGCTGAAGGGCTTTGTTGGTGGGGCAATGACCCCTTGA 7137
QY 178 AAGAGTCCACAGGAAAGAGTCACTCTGTAAACACTTTCAGTTTCATCAAAATGGCCAGAC 237
DB 7138 GATATTCAGACAGGAGGTGGCCATATTTCTCATCTGAATTTCACTGCTGGCCAGAC 7197
QY 238 CATGGCACTTCCTCCTCAGTATGTTTTTTCATCAAAATATGTCGTTATGTGAGGAGAGC 297
DB 7198 CATGATACACCTTCTCAACAGCATGATCTGCTTACTTTTCTCTCTACATGATGACACATC 7257
QY 298 CACATTACAGGACCCCTCCTT 318
DB 7258 CACAGATCAGGCCCAATCAAT 7278

RESULT 5
US-08-854-585-1
Sequence 1, Application US/08854585
Patent No. 6114140
GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas K. and Stman, Arne
TITLE OF INVENTION: Density Enhanced Protein Tyrosine Phosphatase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, Suite 6300
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/854,585
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/237,940
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27866/31954
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5117 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 350..4364
US-08-854-585-1

Query Match 14.8%; Score 47.2; DB 3; Length 5117;
Best Local Similarity 51.1%; Pred. No. 1.8e-05;
Matches 140; Conservative 0; Mismatches 128; Indels 6; Gaps 1;
QY 1 GAAATAATTGTAATGTTATGCTATGATACACAGAGAGATAGAGGTGGAGTTATCAAG 60
Db 3734 GAGAAAAATGATATGCCATCATTTATGCTAAATGTTGTAACAGGGAAGAACAAA 3793
QY 61 TGTTCAGTACTGGCCCGTTCTCTGAAGGAGCGTTTGGAAATTCACACACTTTTCATGTC 120
Db 3794 TGTGAGGAGTATTGGCC-----TCCAAGCAGGCTCAGGACTATGAGACATAAATGTG 3847
QY 121 CTTCTGGAGAACTTTCAGATAACTCAGTATTTGTCTATCCGAATATTTCAAATTTGTGAAG 180
Db 3848 GCAATGACATCAGAAAATGTTCTTCCGGAATGGACCATCAGAGATTTTCACAGTGAATAAT 3907
QY 181 AAGTCCACAGGAAAGTCACTCTGTAAACACATTCAGTTTCATCAATGCGCCAGACCAT 240
Db 3908 ATCCAGACAAGTGAGAGTCACCTCTCAGACAGTTCATTTCCCTCCTCGCCAGACCCAC 3967
QY 241 GGCACCTCCTCAGTAGATTTTTCATCAAT 274
Db 3968 GGTGTTCCCGACACCACTGACCTGCTCATCAACT 4001

RESULT 6
PCT-US95-05512-1
Sequence 1, Application PC/TUS9505512
GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas K. and stman, Arne
TITLE OF INVENTION: Density Enhanced Protein Tyrosine
TITLE OF INVENTION: Phosphatase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 233 South Wacker Drive, Suite 6300
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05512

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27866/31954
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5117 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 350..4364
PCT-US95-05512-1

Query Match 14.8%; Score 47.2; DB 5; Length 5117;
Best Local Similarity 51.1%; Pred. No. 1.8e-05;
Matches 140; Conservative 0; Mismatches 128; Indels 6; Gaps 1;
QY 1 GAAATAATTGTAATGTTATGCTATGATACACAGAGAGATAGAGGTGGAGTTATCAAG 60
Db 3734 GAGAAAAATGATATGCCATCATTTATGCTAAATGTTGTAACAGGGAAGAACAAA 3793
QY 61 TGTTCAGTACTGGCCCGTTCTCTGAAGGAGCGTTTGGAAATTCACACACTTTTCATGTC 120
Db 3794 TGTGAGGAGTATTGGCC-----TCCAAGCAGGCTCAGGACTATGAGACATAAATGTG 3847
QY 121 CTTCTGGAGAACTTTCAGATAACTCAGTATTTGTCTATCCGAATATTTCAAATTTGTGAAG 180
Db 3848 GCAATGACATCAGAAAATGTTCTTCCGGAATGGACCATCAGAGATTTTCACAGTGAATAAT 3907
QY 181 AAGTCCACAGGAAAGTCACTCTGTAAACACATTCAGTTTCATCAATGCGCCAGACCAT 240
Db 3908 ATCCAGACAAGTGAGAGTCACCTCTCAGACAGTTCATTTCCCTCCTCGCCAGACCCAC 3967
QY 241 GGCACCTCCTCAGTAGATTTTTCATCAAT 274
Db 3968 GGTGTTCCCGACACCACTGACCTGCTCATCAACT 4001

RESULT 7
US-09-081-345-1
Sequence 1, Application US/09081345
Patent No. 6228641
GENERAL INFORMATION:
APPLICANT: Bahija Jallal
APPLICANT: Gregory D. Plowman
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: PTP04 RELATED DISORDERS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/081,345
FILING DATE: Herewith

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/047,222
FILING DATE: May 20, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 234/253
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3580 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-081-345-1

Query Match 13.8%; Score 44.2; DB 4; Length 3580;
Best Local Similarity 49.8%; Pred. No. 0.00014;
Matches 143; Conservative 0; Mismatches 138; Indels 6; Gaps 1;

OY 1 GAAATAATTGTAAGTTATTGCTATGATACACAGACATAGAGGTGAGTTATCAAG 60
DB 407 GAATATAGTGTCTTATCATCTGTTATGGCATGCATGAGTATGAAATGGGAAGAAAAAG 466
OY 61 TGTTCAGTTACTGGCCCGCTTCTCTGGAAGGAGCCCTTGGAAATCAACACCTTTCATGTC 120
DB 467 TGTGAGCGTACTGGCTAGCAGGAGAGATGCAGCTGGAATTTGGCCCTTCTCTGTA 526
OY 121 CTCTCGGAGAACTTTCAGATACTCAGTATTTTGTGATCCGGAATATTTCAAAATTTGTAAG 180
DB 527 TCCTGTGAAGCTGAAAAAGGAAATCTGATATATAATCAGGACTCTAAAAGTT----- 580
OY 181 AGTCCAGGAGAAAGTCACTCTGTAAACACACTTGCAGTTTCAATCAATGGCCAGACCAT 240
DB 581 AAGTTCAATAGTGAAGTCACTCTGTAACAGTATCTACAGTATTTCAATCAAGAAATTTGGCCAGACCAT 640
OY 241 GGCACCTCTCGCTCAGTATGATTTTTCATCAATATGTCGGTTATGT 287
DB 641 GATGTACCTTCACTATAGACCCCTATCTTGACCTCATCTGGGATGT 587

RESULT 8
US-08-202-389-3
Sequence 3, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutzky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: BI992-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..285
US-08-202-389-3

Query Match 12.1%; Score 38.8; DB 1; Length 285;
Best Local Similarity 50.0%; Pred. No. 0.0025;
Matches 97; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
OY 93 CTCTGAAGGAGCCCTTGGAAATTCACACACTTTCATGTCTCTCGAGAACTTTTCAGATAA 142
DB 20 CTTCAGAGAGGCTCAGGACTAGGGGACATACTGTGGCAATGACATCAGAACTTTGTC 79
OY 143 CTCAGTATTTGTATCCGGAATATTTCAAAATTTGTAAGAAAGTCCACAGGAAAGAGTCACT 202
DB 80 TTCCGGAATGGACCATCAGAGATTTTGTGTGAAATAATATGCAGAGTAGTGAGAGTCACTC 139
OY 203 CTGTAAACACACTTGCAGTTTCATCAATGGCCAGACCATGCGCACTCTCTGCTCAGTAGATT 262
DB 140 CTCTGCGGAGTTCCATTTCCACCTCTCGCTGACCATGCTGTCTTCTTGACACACCGAC 199
OY 263 TTTTCATCAAAATAT 276
DB 200 TGCTCATCAACTTT 213

RESULT 9
US-08-811-481-30
Sequence 30, Application US/08811481
Patent No. 6300093
GENERAL INFORMATION:
APPLICANT: Kindsvogel, Wayne
APPLICANT: Jelinek, Laura J.
APPLICANT: Sheppard, Paul O.
APPLICANT: Hagopian, William A.
APPLICANT: Lagasse, James M.
TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESS: ZymoGenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

RESULT 13
US-08-239-276-10
; Sequence 10, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/239,276
; FILING DATE: 05-MAY-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-239-276-10

Query Match	11.4%	Score 36.6;	DB 2;	Length 3311;
Best Local Similarity	47.8%	Pred. No. 0.032;		
Matches 139;	Conservative	0;	Mismatches 149;	Indels 3;
				Gaps 1;
QY	10	TGTAATCTTATTGCTATGATAAACAGACAGATACGAAGGTGGAGTTATCAAGTGTGCACT	69	
DB	2170	TGCACGTCATCGTCATGCTGACCCCGTGGTGGAGGATGGTGTCACAGCTGTGACCGC	2229	
QY	70	TACTGGCCGCTTCTCTGGAAGGAGCCCTTTGGAATTCAACACTTTCATGTCTCTCTGGAG	129	
DB	2230	TACTGGCCAGATGAGGTCCTCCCTCTACCACTATATGAGTGAACCTGGTGTGCGAG	2289	
QY	130	AACCTTCAGATAACTCAGTATTTTGTATCGGAATATTTCAAATTTGTCAAGAAGTCCACA	189	
DB	2290	CACATCTGG---TCCGAGGACTTTCTGGTGGGAGCTTCTACCTGAAGAAGCTGCAGACC	2346	
QY	190	GGAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAATGGCCAGACCATGGCACTCCT	249	
DB	2347	CAGGACGGCGACGCTCACACAGTTCACATTCCTCTACGTGGCCGGCAGAGGGCACCGC	2406	

Qy 250 GCCTCAGTAGATTTTTTCATCAAAATATGTCGGTTATGTGAGGAAGGCCAC 300
||||| I | I | I | I | I | I | I | I | I | I | I | I | I | I | I |
Db 2407 GCCTCCACGGCGCCCTGCTGGACTTCCGCAGGAAGTGAAACAAGTGTCTAC 2457

RESULT 14

US-08-468-579B-10
Sequence 10, Application US/08468579B
Patent No. 5981700
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-458-579B-10

```

Query Match          11.4%; Score 36.6; DB 2; Length 3311;
Best Local Similarity 47.8%; Pred. No. 0.032;
Matches 139; Conservative 0; Mismatches 149; Indels 3; Gaps 1;

Qy . 10 TGTAAATGTTCTATGATTAACGACAGAGATAGAAGGTGGAGTTATCAAGTGTTCAGT 69
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2170 TGCACCGCACATGCTATGCTACCCCGCTGGTGGAGGATGGTGAACGACAGTGTACCGC 2229

Qy 70 TACTGGCCGGTTTCTCTGAAGAGGCGTTTGGAAATTAACAACACTTTCATGTCTCTCTGGAG 129
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2230 TACTGGCCAGATGAGGGTGCCTCCCTCTACCAACGCTATATGAGGTGAACCTGGTGTCCGAG 2289

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:16:28 ; Search time 27.53 Seconds
(without alignments)
328.258 Million cell updates/sec

Title: US-09-095-478A-8

Perfect score: 122

Sequence: 1 DFQGMWNNCNIAMITRE.....VRKSHITGPLLVHCTAGVGR 122

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_1101.*

- 1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	100.0	122	20	Rat PTP10. Rattus
2	23	18.9	405	20	Mouse PTP05 isofor
3	23	18.9	426	20	Mouse PTP05. Mus
4	23	18.9	463	20	Mouse PTP05 isofor
5	7	5.7	186	20	Wheat Atp-phosphor
6	7	5.7	243	22	C glutamicum prote
7	7	5.7	257	22	Drosophila protein
8	7	5.7	436	20	Protein involved i
9	7	5.7	488	20	A RNA-dependent am
10	7	5.7	488	20	Glutamyl-tRNA(Gln)
11	6	4.9	9	14	Peptide for treati

12	6	4.9	9	14	AAR39296	Endotoxin lipid A
13	6	4.9	9	16	AAW71780	Peptide neutralisi
14	6	4.9	9	18	AAW21627	Antibiotic potenti
15	6	4.9	9	18	AAW21597	Antibiotic potenti
16	6	4.9	9	21	AAW56909	Peptide contained
17	6	4.9	9	21	AAW56939	Peptide contained
18	6	4.9	10	22	AAG86608	Saccharomyces cere
19	6	4.9	10	22	AAG86610	Saccharomyces cere
20	6	4.9	48	17	AAW06711	Helicobacter-speci
21	6	4.9	61	22	AAW17794	Peptide #4228 enco
22	6	4.9	61	22	AAW30304	Peptide #4341 enco
23	6	4.9	61	22	AAW05441	Peptide #4123 enco
24	6	4.9	71	21	AAW53851	Human colon cancer
25	6	4.9	73	21	AAW69156	Peptide HH2040-BF0
26	6	4.9	74	19	AAW41091	Chicken matrix met
27	6	4.9	100	21	AAW02295	Human secreted pro
28	6	4.9	105	21	AAW90691	Human spectrin ple
29	6	4.9	108	19	AAW41092	Chicken matrix met
30	6	4.9	111	22	AAW17183	Peptide #3617 enco
31	6	4.9	111	22	AAW29676	Peptide #3713 enco
32	6	4.9	111	22	AAW04878	Peptide #3560 enco
33	6	4.9	117	19	AAW48090	Aeromonas caviae p
34	6	4.9	127	21	AAW58230	Feline mature GMCS
35	6	4.9	141	21	AAW57566	Arabidopsis thalia
36	6	4.9	142	21	AAW03889	Human secreted pro
37	6	4.9	144	21	AAW58229	Feline GMCSF. Fei
38	6	4.9	144	22	AAW67463	Amino acid sequenc
39	6	4.9	144	22	AAW67464	Amino acid sequenc
40	6	4.9	159	22	AAW39275	Human polypeptide
41	6	4.9	159	22	AAW82299	Human secreted pro
42	6	4.9	179	20	AAW36002	Extended human sec
43	6	4.9	179	22	AAW39714	Human polypeptide
44	6	4.9	185	19	AAW98456	H. pylori GHPO 654
45	6	4.9	193	19	AAW41090	Chicken matrix met

ALIGNMENTS

RESULT 1
AAW89252
ID AAW89252 standard; Protein; 122 AA.
XX
AC AAW89252;
XX
DT 10-MAR-1999 (first entry)
DE Rat PTP10.
XX
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
OS Rattus sp.
XX
PN WO9849317-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08439.
XX
PR 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
PA (SJGE-) SUGEN INC.
XX
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;

```

XX WPI; 1999-009434/01.
DR N-PSDB; AAV81747.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents rat PTP10. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 122 AA;
SQ

Query Match 100.0%; Score 122; DB 20; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.8e-125;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DFWGMMWNNCNCVIAITREIEGGVVKCCSWPVSLKEPLEKFHFVLLNFQITQYFVI 60
   |||||||
Db 1 dfwgmmwennncnviamitreieggvikccswpvsllkeplekfhfvllefnfqtgyfvi 60

QY 61 RIFQIVKSTGKSHSVKHLQFIKWPDHGCTPASVDFFIKYVYVRKSHITGPLLHVCTAGV 120
   |||||||
Db 61 rifqivkstgkshsvkhlqfkwphdgtpasvdffikyvrvrkshitgpllvhctagv 120

QY 121 GR 122
   ||
Db 121 gr 122

RESULT 2
AAW89251
ID AAW89251 standard; Protein; 405 AA.
XX
XX AAW89251;
AC
XX
XX 10-MAR-1999 (first entry)
DT
XX
XX Mouse PTP05 isoform #2.
DE
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
OS
XX
XX WO9849317-A2.
PN
XX
XX 05-NOV-1998.
PD
XX
XX 27-APR-1998; 98WO-US08439.
PF
XX
XX 23-OCT-1997; 97US-0063595.
PR
XX 28-APR-1997; 97US-0044428.
PR
XX 20-MAY-1997; 97US-0047222.
PR

```

```

PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGB-) SUCEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;
PI
XX
XX WPI; 1999-009434/01.
DR N-PSDB; AAV81746.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 158-160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 405 AA;
SQ

Query Match 18.9%; Score 23; DB 20; Length 405;
Best Local Similarity 100.0%; Pred. No. 9.3e-17;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DFFIKYVYVRKSHITGPLLHVHC 116
   |||||||
Db 337 dffikyrvyrkshitgpllvhc 359

RESULT 3
AAW89249
ID AAW89249 standard; Protein; 426 AA.
XX
XX AAW89249;
AC
XX
XX 10-MAR-1999 (first entry)
DT
XX
XX Mouse PTP05.
DE
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
OS
XX
XX WO9849317-A2.
PN
XX
XX 05-NOV-1998.
PD
XX
XX 27-APR-1998; 98WO-US08439.
PF
XX
XX 23-OCT-1997; 97US-0063595.
PR
XX 28-APR-1997; 97US-0044428.
PR
XX 20-MAY-1997; 97US-0047222.
PR

```


PI Fujimori K, Mizutani M, Ohta D;
 DR WPI; 1999-153320/13.
 DR N-PSDB; AAX24355.
 XX
 XX New plant ATP-phosphoribosyl transferase (APRT) encoding DNA
 PT useful for providing herbicide-resistant plants
 XX
 XX Claim 2; Page 34; 58pp; English.
 XX
 XX This is the amino acid sequence of an ATP phosphoribosyl transferase
 CC (APRT) polypeptide of wheat, as encoded by a cDNA clone (see
 CC AAX24355) isolated from a 7-day-old wheat seedling cDNA library.
 CC Another wheat APRT polypeptide (see AAW97852) is encoded by another
 CC isolated cDNA clone (see AAX24354). The isolation of 2 DNA sequences
 CC coding for individual proteins is probably due to the presence of
 CC multiple isoforms encoded by different genes in the wheat genome.
 CC Arabidopsis APRT polypeptides (see AAW97854 and AAW97855) are also
 CC provided. Sequence comparison suggests the wheat polypeptides
 CC lack their N-terminal portions including a signal peptide and
 CC chloroplast transit peptide found in the Arabidopsis APRT proteins.
 CC Using the information obtained from these sequences, APRT enzymes
 CC can be obtained from any plant source. The invention also
 CC encompasses the recombinant production of APRT, probes, and methods
 CC for detecting the presence and form of the APRT gene and
 CC quantitating levels of APRT transcripts in an organism. Expression
 CC cassettes and recombinant vectors are also provided, as well as the
 CC preparation of transgenic plants, plant tissue and seed which has
 CC been stably transformed with a recombinant DNA molecule comprising
 CC a plant promoter and a gene encoding APRT. A DNA molecule encoding
 CC a modified APRT, having at least 1 amino acid modification and
 CC which is tolerant to a herbicide can be used to confer herbicide
 CC tolerance to a plant. Recombinant APRT enzyme can be used to
 CC screen known herbicidal chemicals to determine if they inhibit
 CC APRT, and to screen for candidate herbicides.
 XX
 SQ Sequence 186 AA;

Query Match 5.7%; Score 7; DB 20; Length 186;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 FIEGVI 26
 |||||
 Db 40 eieggi 46

RESULT 6
 AAG91426
 ID AAG91426 standard; Protein; 243 AA.
 XX
 AC AAG91426;
 XX
 XX 26-SEP-2001 (first entry)
 DT
 XX
 DE C glutamicum protein fragment SEQ ID NO: 5180.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PE 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX

PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH66645.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX
 XX Claim 17; SEQ ID NO: 5180; 246pp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 243 AA;

Query Match 5.7%; Score 7; DB 22; Length 243;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 TPASVDF 95
 |||||
 Db 165 tpsvdf 171

RESULT 7
 AAB59369
 ID AAB59369 standard; Protein; 257 AA.
 XX
 AC AAB59369;
 XX
 XX 21-MAR-2001 (first entry)
 DT
 XX
 DE Drosophila protein tyrosine phosphatase #1.
 XX
 KW Protein tyrosine phosphatase; human; mouse; fruit fly; PTP;
 KW substrate trapping.
 XX
 OS Drosophila sp.
 XX
 PN WC200075339-A1.
 XX
 PD 14-DEC-2000.
 XX
 PE 24-MAY-2000; 2000WO-US14211.
 XX
 PR 03-JUN-1999; 99US-0137319.
 PR 16-JUN-1999; 99US-0334575.
 XX
 XX (COLD-) COLD SPRING HARBOR LAB.
 XX
 XX Tonks NK, Zhang S;
 PI WPI; 2001-080598/09.
 DR
 DR New substrate trapping mutant protein tyrosine phosphatases (PTP) in
 PT which the wild type PTP catalytic domain invariant aspartate is

```

Query: Match
Best Local Similarity 5.7%; Score 7; DB 20; Length 436;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 451 NFQITQY 57
Db 96 nfqltqy 102
RESULT: 9

```

AD
XX
ADW97397 standard; Protein; 488 AA.
XX
AC
AA097397;
XX
DT
17-MAY-1999 (first entry)
XX
XX
A RNA-dependent amino transferase (ratB) subunit.
DE
XX

KW	protein synthesis; classic ocular trachoma; inclusion conjunctivitis;
KW	gepital trachoma; infant pneumonitis; Lymphogranuloma Venereum;
KW	incipient trachoma; keratitis; papillary hypertrophy;
KW	corneal infiltration; vulvovaginitis; ear infection; prostaticitis;
KW	climatic bubo; bacterial infection; sexually transmitted disease;
KW	fertility; bacterial adhesion; matrix protein; wound; body implant.
XX	
XX	Glamydia trachomatis.
XX	
XX	EP899339-Al.
XX	
XX	03-MAR-1999.
PD	
XX	
XX	04-AUG-1998; 98EP-0306205.
XX	
XX	13-AUG-1997; 97US-0910313.
XX	
XX	(SMIK) SMITHKLINE BEECHAM CORP.
XX	
XX	

XX	13.
XX	WPI; 1999-144810/13.
DR	N-PSDB; AAX16047.
DR	
XX	
XX	
XX	New Chlamydia trachomatis RNA-dependant Amino Transferase (ratB)
PT	subunit gnc and protein - useful as diagnostic reagents and for
PT	prevention and treatment of Chlamydia infections, which cause
PT	sexually transmitted diseases and infant pneumonitis
XX	
XX	Claim 6; Page 7; 26pp; English.
PS	
XX	
XX	The present sequence represents a RNA-dependent amino transferase (ratB)
CC	subunit polypeptide. The protein catalyses the formation of Gln-tRNA Gln,
CC	which is required for accurate protein synthesis in prokaryotes. RatB
CC	polypeptides and polynucleotides are useful for diagnosing diseases due
CC	to an infection of an organism with the ratB gene, by determining the
CC	presence of the nucleic acid encoding ratB, and/or analysing for the

presence or amount of ratB polypeptide in the sample, they can diagnose the stage and type of infection. RatB polypeptides are also useful for screening for compounds which affect activity of the protein. These compounds can be used in treatment to inhibit (antagonist i.e. antibacterial drugs) or enhance (agonist) ratB activity. In addition to direct administration of ratB polypeptides to treat conditions associated with a lack of ratB polypeptide, or direct administration of antisense sequences to prevent expression. RatB polypeptides and antibodies induce an immune response to immunise and prevent disease. Diseases diagnosed, prevented or treated include: classic ocular trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis, lymphogranuloma venereum, incipient trachoma, keratitis, papillary hypertrophy, corneal

CC infection, vulvovaginitis, ear infection, proctitis, and climatic
CC bubo; and bacterial infections, especially chlamydia trachomatis

CC RatB polypeptides, polynucleotides and their (ant)agonists can prevent
 CC adhesion of bacteria to matrix proteins, and are useful for use on wounds
 CC and body implants to prevent bacterial infection.

XX Sequence 488 AA;

Query Match 5.7%; Score 7; DB 20; Length 488;

Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57
 Db 94 nfgitqy 100

RESULT 10

AAW78471
 ID AAW78471 standard; Protein: 488 AA.

XX AC AAW78471;

XX DT 11-MAY-1999 (first entry)

DE Glutamyl-tRNA(Gln) amidotransferase subunit ratB subunit.

XX Glutamyl-tRNA(Gln) amidotransferase; ratB; diagnosis; trachoma; adhesion;
 KW Conjunctivitis; infant pneumonitis; Lymphogranuloma Venereum; keratitis;
 KW papillary hypertrophy; corneal infiltration; vulvovaginitis; prostaticitis;
 KW ear infection; climatic bubo; infection; sexually transmitted disease;
 KW infertility; antagonist; matrix protein; wound; body implant.

XX Chlamydia trachomatis.

XX EP892061-A2.

XX PD 20-JAN-1999.

XX PF 01-JUL-1998; 98EP-0305247.

XX PR 18-JUL-1997; 97US-0896342.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Black MT, Reichard R;

XX WI: 1999-083578/08.

DR N-PSDB; AAX17999.

XX New Chlamydia trachomatis polypeptide and polynucleotide - useful
 PT as diagnostic reagents and for prevention and treatment of diseases
 PT and sexually transmitted diseases and infertility caused by
 PT Chlamydia trachomatis infections

XX Claim 12; Page 6; 23pp; English.

XX This sequence represents the Chlamydia trachomatis glutamyl-tRNA(Gln)
 CC amidotransferase subunit ratB protein sequence. RatB polypeptides and
 CC polynucleotides are useful for diagnosing diseases e.g. classic ocular
 CC trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis,
 CC Lymphogranuloma Venereum, incipient trachoma, keratitis, papillary
 CC hypertrophy, corneal infiltration, vulvovaginitis, ear infection,
 CC prostaticitis, and climatic bubo, and bacterial infections, especially
 CC Chlamydia trachomatis infections which cause sexually transmitted
 CC diseases and infertility. RatB polypeptides, polynucleotides and their
 CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and
 CC are useful for use on wounds and body implants to prevent bacterial
 CC infection.

XX Sequence 488 AA;

Query Match 5.7%; Score 7; DB 20; Length 488;

Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57
 Db 94 nfgitqy 100

RESULT 11

AAAR33534
 ID AAR33534 standard; peptide: 9 AA.

XX AC AAR33534;

XX DT 07-JUL-1993 (first entry)

DE Peptide for treating septic shock.

XX Toxic shock; blood endotoxin removal; serum; diagnostic reagent;
 KW cytokine release control; treatment; pertussis; bacterial meningitis;
 KW HIV related infections; polymyxin B.

OS Synthetic.

XX ZA9200943-A.

XX PD 25-NOV-1992.

XX PF 10-FEB-1992; 92ZA-0000943.

XX PR 11-FEB-1991; 91US-0658744.

XX (PORR/) PORRO M.

XX PI Porro M;

XX WI: 1993-094304/11.

XX New peptide for treatment or prevention of toxic shock - comprises
 PT specified sequences of aminoacid(s) and analogs
 PT comprising sequences retro-orientated

XX Claim 12; Page 32; 39pp; English.

XX This peptide is a specific example of a generic peptide of
 CC formula R-(Lys/Arg/His - Phe/Tyr/Trp - Leu/Ile/Val)n-R, where
 CC n = 1-100 and each R is H, an amino acid residue or a fatty acid
 CC residue. The peptide is useful for treating or preventing septic
 CC shock, mixing with polymyxin B to reduce its toxicity; removing
 CC endotoxins from blood, sera or other fluids (in vivo or in
 CC vitro); controlling release of cytokines induced by endotoxins;
 CC as diagnostic reagents to detect and quantify toxins in blood
 CC or sera; preparing non-toxic antigenic complexes of lipid A or
 CC lipopolysaccharide (LPS); and for treating pertussis, bacterial
 CC meningitis and HIV-related infections. The usual dose is 10-100
 CC µg/kg/day, given parenterally. It binds to the same sites as
 CC polymyxin B, i.e. it inhibits all the toxic effects of lipid A. It
 CC has no antibiotic activity; does not lyse erythrocytes; has no
 CC toxicity in mice when injected at 50mg/kg and is relatively unstable
 CC against proteases.

XX Sequence 9 AA;

Query Match 4.9%; Score 6; DB 14; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YRVYVR 104
 Db 2 yrvyvr 7

RESULT 12

AAR39296
ID AAR39296 standard; peptide; 9 AA.
XX
AC
XX
AC AAR39296;
XX
DT 22-DEC-1993 (first entry)
XX
DE Endotoxin lipid A neutralising peptide.
XX
KW Bacterial endotoxin; non-toxic; treatment; prevention; toxic shock;
KW polymyxin B; PMB; toxicity reduction; reduce; blood endotoxin;
KW removal; sera; vaccines; drug solns.; solutions; cytokine release;
KW control; in vivo; in vitro; detoxification; detection;
KW quantification.
XX
OS Synthetic.
XX
OS
XX
PN WO9314115-A.
XX
PD 22-JUL-1993.
XX
PF 14-MAY-1992; 92WO-EP01060.
XX
PR 16-JAN-1992; 92US-0819893.
XX
PA (PORR/) PORRO M.
XX
PI Porro M;
XX
XX
XX WPI; 1993-243143/30.
DR
XX
XX New peptide(s) which neutralise lipid A of bacterial endotoxin -
PT forming non-toxic, antigenic complex, used to treat or prevent
PT septic shock, in vaccines to detoxify blood, etc.
XX
PS Claim 12; Page 33; 45pp; English.
XX

The sequence is that of a peptide which binds to the lipid A
CC component of bacterial endotoxin at the same site as polymyxin B
CC (PMB) and with about the same affinity to produce a non-toxic
CC antigenic complex. Unlike PMB it is not toxic, is susceptible to
CC proteolytic degradation in serum, has no antibiotic activity and no
CC haemolytic action. It is especially used to treat or prevent septic
CC shock; to reduce toxicity of PMB; to remove endotoxins from blood,
CC sera, vaccines, drug solns. etc.; to control release of cytokines
CC induced by endotoxins; for in vivo or in vitro detoxification of
CC bacterial endotoxins, and to detect or quantify endotoxins in blood
CC products.
XX
SO Sequence 9 AA;

Query Match 4.9%; Score 6; DB 14; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 YVRVVR 104
Db 2 yvryvr 7
|||||

RESULT 13

AAR71780
ID AAR71780 standard; peptide; 9 AA.
XX
AC AAR71780;
XX
DT 01-OCT-1995 (first entry)
XX
DE Peptide neutralising toxicity of Lipid A.
XX
KW Endotoxin; Lipid A; septic shock; lipopolysaccharide.

XX Synthetic.
OS
XX
XX WO9503327-A.
XX
XX 02-FEB-1995.
XX
XX 21-JUL-1994; 94WO-EP02413.
XX
XX 26-JUL-1993; 93US-0097830.
XX
XX (BIOS-) BIOSYNTH SRL.
XX
XX Porro M;
XX
XX WPI; 1995-075190/10.
XX
XX New peptide(s) for neutralising LPS endotoxin - comprising
PT repeating units of a basic amino acid or basic and hydrophobic
PT amino acids
XX
XX Claim 11; Page 20; 26pp; English.

New peptides are claimed which are linear or cyclic peptides of formula:
CC (A)n, where A is the cationic amino acid Lys or Arg and n is 7-10;
CC (AB)m, where A is Lys or Arg, B is the hydrophobic amino acid Val, Leu,
CC Ile, Tyr, Phe or Trp, and m is 3 or greater; and (ABC)p, where A is
CC Lys or Arg, B and C are Val, Leu, Ile, Tyr, Phe or Trp, and p is 2 or
CC greater.
CC The peptides bind to Lipid A of endotoxins and do not exhibit haemolytic
CC activity. Hence they can be used therapeutically to treat septic shock
CC and also in vitro to detoxify vaccines, drug solutions, injectable
CC nutrient solutions, etc.
CC The present sequence is a specifically claimed example of the new
CC peptides.
XX
SO Sequence 9 AA;

Query Match 4.9%; Score 6; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 YVRVVR 104
Db 2 yvryvr 7
|||||

RESULT 14

AAW21627
ID AAW21627 standard; peptide; 9 AA.
XX
AC AAW21627;
XX
XX 26-AUG-1997 (first entry)
XX
XX Antibiotic potentiating peptide #39.
DE Potentiate; antibiotic; microbial infection; lipopolysaccharide;
KW permeability; outer bacterial membrane.
XX
XX Synthetic.
XX
XX WO9638163-A1.
XX
XX 05-DEC-1996.
XX
XX 29-MAY-1996; 96WO-EP02313.
XX
XX 31-MAY-1995; 95US-0456112.
XX
XX (BIOS-) BIOSYNTH SRL.

PI Porro M, Varra M;
XX WPI; 1997-034095/03.
XX
XX Potentiating activity of antibiotic with peptide contg. cationic
PT amino acid sequence - reduces dose of antibiotic required
XX
XX Claim 43; Page 28; 37pp; English.
PS
XX The sequences given in AAW21589-633 represent peptides which act to
CC potentiate the activity of an antibiotic when they are co-administered
CC with the antibiotic. Compositions containing these peptides are used
CC to treat or prevent microbial infections. These peptides bind to
CC lipopolysaccharide on the bacteria so may increase permeability of
CC the outer bacterial membrane to the antibiotic, allowing a reduction
CC in the dose of antibiotic required by 10-90% of the normal dose for
CC in vivo or in vitro application. Any toxic side effects are
CC correspondingly reduced.
XX
XX Sequence 9 AA;
SQ

Query Match 4.9%; Score 6; DB 18; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 99 YRYVR 104
Db |||||
2 yryvr 7

RESULT 15
AAW21597
ID AAW21597 standard; peptide; 9 AA.
XX
XX AAW21597;
XX
XX 26-AUG-1997 (first entry)
DE Antibiotic potentiating peptide #9.
XX
XX Potentiate; antibiotic; microbial infection; lipopolysaccharide;
KW permeability; outer bacterial membrane.
XX
XX Synthetic.
OS
XX WO9638163-A1.
PN
XX
XX 05-DEC-1996.
PD
XX 29-MAY-1996; 96WO-EP02313.
PF
XX 31-MAY-1995; 95US-0456112.
PR
XX (BIOS-) BIOSYNTH SRL.
XX
XX PA
XX
XX PI Porro M, Varra M;
XX
XX WPI; 1997-034095/03.
DR
XX
XX Potentiating activity of antibiotic with peptide contg. cationic
PT amino acid sequence - reduces dose of antibiotic required
XX
XX Claim 13; Page 24; 37pp; English.
PS
XX The sequences given in AAW21589-633 represent peptides which act to
CC potentiate the activity of an antibiotic when they are co-administered
CC with the antibiotic. Compositions containing these peptides are used
CC to treat or prevent microbial infections. These peptides bind to
CC lipopolysaccharide on the bacteria so may increase permeability of
CC the outer bacterial membrane to the antibiotic, allowing a reduction
CC in the dose of antibiotic required by 10-90% of the normal dose for
CC in vivo or in vitro application. Any toxic side effects are

CC correspondingly reduced.
XX
XX Sequence 9 AA;
SQ

Query Match 4.9%; Score 6; DB 18; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 99 YRYVR 104
Db |||||
2 yryvr 7

Search completed: March 30, 2002, 08:26:44
Job time: 616 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:22:33 ; Search time 15.23 seconds
(without alignments)
180.263 Million cell updates/sec

Title: US-09-095-478A-8
Perfect score: 122
Sequence: 1 DFGMMWNNCNIAMITRE.....VRKSHITGPLLHICTAGVGR 122

Scoring table:
OLIGO
Capex 60.0 , Capex 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	5.7	257	2	US-08-685-992-6
2	7	5.7	257	2	US-09-144-925-6
3	7	5.7	488	4	US-08-910-313-2
4	6	4.9	9	1	US-07-819-893-10
5	6	4.9	9	1	US-08-280-397-10
6	6	4.9	9	1	US-08-097-830E-9
7	6	4.9	9	2	US-08-456-112B-9
8	6	4.9	9	2	US-08-456-112B-39
9	6	4.9	48	2	US-08-849-480A-8
10	6	4.9	105	3	US-08-407-165-3
11	6	4.9	117	2	US-08-910-856-4
12	6	4.9	260	4	US-08-957-130-15
13	6	4.9	272	2	US-08-446-345-34
14	6	4.9	327	2	US-08-907-674-3
15	6	4.9	327	2	US-09-215-087-3
16	6	4.9	327	3	US-09-391-959-3
17	6	4.9	331	2	US-08-907-674-1
18	6	4.9	331	2	US-09-215-087-1
19	6	4.9	331	3	US-09-391-959-1
20	6	4.9	440	1	US-08-061-062A-6
21	6	4.9	440	1	US-08-061-062A-8
22	6	4.9	440	3	US-08-536-150-6
23	6	4.9	440	3	US-08-536-150-8
24	6	4.9	579	4	US-08-743-168B-36
25	6	4.9	581	4	US-08-743-168B-40
26	6	4.9	590	4	US-08-743-168B-43
27	6	4.9	1174	2	US-08-446-345-36

28 6 4.9 1375 3 US-08-665-259-26 Sequence 26, Appl
29 6 4.9 1375 3 US-08-762-500-26 Sequence 26, Appl
30 2548 4 US-09-172-422-1 Sequence 1, Appl
31 5 4.1 7 2 US-08-292-968-31 Sequence 31, Appl
32 5 4.1 7 2 US-08-467-974-31 Sequence 31, Appl
33 5 4.1 7 2 US-08-467-974-31 Sequence 31, Appl
34 5 4.1 7 2 US-08-467-976-31 Sequence 31, Appl
35 5 4.1 7 4 US-09-082-514-31 Sequence 12, Appl
36 5 4.1 8 2 US-08-342-930-12 Sequence 4, Appl
37 5 4.1 8 3 US-08-336-553A-4 Sequence 33, Appl
38 5 4.1 8 3 US-08-336-553A-55 Sequence 55, Appl
39 5 4.1 8 3 US-08-336-553A-74 Sequence 74, Appl
40 5 4.1 8 4 US-08-916-935-10 Sequence 10, Appl
41 5 4.1 9 3 US-07-792-600-34 Sequence 34, Appl
42 5 4.1 9 3 US-09-157-021-34 Sequence 34, Appl
43 5 4.1 9 3 US-09-156-842-34 Sequence 34, Appl
44 5 4.1 9 4 US-09-169-015-14 Sequence 14, Appl
45 5 4.1 9 4 US-09-169-015-14 Sequence 14, Appl

ALIGNMENTS

RESULT , 1
US-08-685-992-6
: Sequence 6, Application US/08685992
: Patent No. 5912138
: GENERAL INFORMATION:
: APPLICANT: Tonks, Nicholas
: APPLICANT: Flint, Andrew J.
: TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
: TITLE OF INVENTION: TYROSINE PHOSPHATASES
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
: STREET: Two Millia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows 95
: SOFTWARE: FASTSEQ for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/685,992
: FILING DATE: 25-JUL-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: CSHL96-03
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 781-861-6240
: TELEFAX: 781-861-9540
: TELEX:
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 257 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-08-685-992-6

Query Match 5.7%; Score 7; DB 2; Length 257;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGT 90
| | | | |
Db 156 WPDHGT 162

RESULT 2

US-09-144-925-6
; Sequence 6, Application US/09144925
; Patent No. 5951979
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas
; APPLICANT: Flint, Andrew J.
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02421-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/144,925
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/685,992
; FILING DATE: July 25, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL96-03Z
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: peptide
US-09-144-925-6

Query Match 5.7%; Score 7; DB 2; Length 257;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGT 90
| | | | |
Db 156 WPDHGT 162

RESULT 3

US-08-910-313-2
; Sequence 2, Application US/08910313
; Patent No. 6171838
; GENERAL INFORMATION:
; APPLICANT: Black, Michael Terence
; TITLE OF INVENTION: No. 6171838el RatB
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,313
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10072
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 488 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-910-313-2

Query Match 5.7%; Score 7; DB 4; Length 488;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57
| | | | |
Db 94 NFOITQY 100

RESULT 4

US-07-819-893-10
; Sequence 10, Application US/07819893
; Patent No. 5371186
; GENERAL INFORMATION:
; APPLICANT: Porro, Massimo
; TITLE OF INVENTION: Synthetic Peptides for Detoxification
; TITLE OF INVENTION: Of Bacterial Endotoxins and for the
; TITLE OF INVENTION: Prevention and Treatment of Septic
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hedman, Gibson, Costigan & Hoare
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/819,893
; FILING DATE: 19920115
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: AMINO ACIDS
TOPOLOGY: linear
US-07-819-893-10

Query Match 4.9%; Score 6; DB 1; Length 9;
Best Local Similarity 100.08; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVYVR 104
Db 2 YVYVR 7

RESULT 5
US-08-280-397-10
Sequence 10, Application US/08280397
Patent No. 5589459
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: Synthetic Peptides for Detoxification
TITLE OF INVENTION: of Bacterial Endotoxins and for the
TITLE OF INVENTION: Prevention and Treatment of Septic
TITLE OF INVENTION: Shock
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan, P.C.
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/280,397
FILING DATE: 07/26/94
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/819,893
FILING DATE: 01/16/92
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-002A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acids
TOPOLOGY: linear
US-08-280-397-10

Query Match 4.9%; Score 6; DB 1; Length 9;
Best Local Similarity 100.08; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVYVR 104

Db 2 YVYVR 7

RESULT 6
US-08-097-830E-9
Sequence 9, Application US/08097830E
Patent No. 5652211
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: Peptides For Neutralizing The
TITLE OF INVENTION: Toxicity of Lipid A
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan, P.C.
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/097,830E
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-097-830E-9

Query Match 4.9%; Score 6; DB 1; Length 9;
Best Local Similarity 100.08; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVYVR 104
Db 2 YVYVR 7

RESULT 7
US-08-456-112B-9
Sequence 9, Application US/08456112B
Patent No. 5834430
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: POTENTIATION OF ANTIBIOTICS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage

COMPUTER: LEADING EDGE 486
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,112B
FILING DATE: May 31, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-08-456-112B-9

Query Match 4.9%; Score 6; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YRYVR 104
Db 2 YRYVR 7

RESULT 8
US-08-456-112B-39
Sequence 39, Application US/08456112B
Patent No. 5834430
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: POTENTIATION OF ANTIBIOTICS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: LEADING EDGE 486
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,112B
FILING DATE: May 31, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: circular

US-08-456-112B-39

Query Match 4.9%; Score 6; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YRYVR 104
Db 2 YRYVR 7

RESULT 9
US-08-849-480A-8
Sequence 8, Application US/08849480A
Patent No. 5981184
GENERAL INFORMATION:
APPLICANT: MELCHERS, Klaus
TITLE OF INVENTION: SCREENING MODEL
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: 400 - 7th Street, N. W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/849,480A
FILING DATE: 02-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04711
FILING DATE: 30-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P4442970.3
FILING DATE: 02-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19505645.0
FILING DATE: 18-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: AISENBERG, Irwin M.
REGISTRATION NUMBER: 19,007
REFERENCE/DOCKET NUMBER: 8125/P60984US0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/638-6666
TELEFAX: 202/393-5350
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
STRAIN: Helicobacter pylori 69A
INDIVIDUAL ISOLATE: Clinical isolate 69A
IMMEDIATE SOURCE:
LIBRARY: Helicobacter pylori 69A - gene library in
LIBRARY: vector pRH160
CLONE: pRH948
US-08-849-480A-8

Query Match 4.9%; Score 6; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 16;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 SVDFFI 97
|||||
Db 9 SVDFFI 14

RESULT 10
US-08-407-165-3
; Sequence 3, Application US/08407165
; Patent No. 6054280
; GENERAL INFORMATION:
; APPLICANT: LEMMON, MARK A.
; APPLICANT: FERGUSON, KATHRYN M.
; APPLICANT: SIGLER, PAUL B.
; APPLICANT: SCHLESSINGER, JOSEPH
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF
; FILE OF INVENTION: PH DOMAIN SIGNAL TRANSDUCTION DISORDERS
; FILE REFERENCE: 211/156
; CURRENT APPLICATION NUMBER: US/08/407,165
; CURRENT FILING DATE: 1995-03-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Spectrin
US-08-407-165-3

Query Match 4.9%; Score 6; DB 3; Length 105;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 PVSKE 38
|||||
Db 55 PVSKE 60

RESULT 11
US-08-910-856-4
; Sequence 4, Application US/08910856
; Patent No. 5981257
; GENERAL INFORMATION:
; APPLICANT: FUKUI, TOSHIAKI
; APPLICANT: DOI, YOSHIHARU
; TITLE OF INVENTION: POLYESTER SYNTHASE GENE AND PROCESS
; FILE OF INVENTION: FOR PRODUCING POLYESTER
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 4225 EXECUTIVE SQUARE, SUITE 1400
; CITY: LA JOLLA
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,856
; FILING DATE: 13-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 214509/1996
; FILING DATE: 14-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 199979/1997
; FILING DATE: 25-JUL-1997

ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07898/016001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-856-4

Query Match 4.9%; Score 6; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KKSTGK 72
|||||
Db 111 KKSTGK 116

RESULT 12
US-08-957-130-15
; Sequence 15, Application US/08957130
; Patent No. 6290959
; GENERAL INFORMATION:
; APPLICANT: WU, Xue-Ru
; APPLICANT: SUN, Tung-Tien
; TITLE OF INVENTION: METHOD FOR SCREENING COMPOUNDS FOR
; FILE OF INVENTION: INHIBITING BACTERIAL ATTACHMENT TO HOST CELL RECEPTORS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,130
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: SUN-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 260 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-130-15

Query Match 4.9%; Score 6; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 SLKEPL 40

Db 195 SLKEPL 200

RESULT 13

```
US-08-446-345-34
; Sequence 34, Application US/08446345
; Patent No. 5831009
; GENERAL INFORMATION:
; APPLICANT: Ulrich, Axel
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASES PTP-D1
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-446-345-34
```

Query Match 4.9%; Score 6; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 PLLVHC 116
Db 201 PLLVHC 206

RESULT 14

```
US-08-907-674-3
; Sequence 3, Application US/08907674
; Patent No. 5919685
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/907,674
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 433611
; US-08-907-674-3
```

Query Match 4.9%; Score 6; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 PDHGTG 90
Db 111 PDHGTG 116

RESULT 15

```
US-09-215-087-3
; Sequence 3, Application US/09215087
; Patent No. 5981244
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215.087
; FILING DATE:
```

CLASSIFICATION:
PRIOR APPLICATION DATA: 08/907,674
APPLICATION NUMBER: 08/907,674
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 433611
US-09-215-087-3

Query Match 4.9%; Score 6; OB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 PDHCTP 90
|||||
Db 111 PONGTP 116

Search completed: March 30, 2002, 08:27:44
Job time: 311 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:17:58 ; Search time 19.18 seconds
(without alignments)
484.531 Million cell updates/sec

Title: US-09-095-478A-8

Perfect score: 122

Sequence: 1 DFWGMMWNNCNVIAMITRE.....VRKSHITGILLVHCTAGVGR 122

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	8.2	398	2 T08716	protein-tyrosine-p
2	7	5.7	310	2 T22013	hypothetical prote
3	7	5.7	377	1 A48711	protein-tyrosine-p
4	7	5.7	397	2 T21154	hypothetical prote
5	7	5.7	488	2 G71568	probable (pet112)
6	7	5.7	488	2 B81722	glutamyl-tRNA(Gln)
7	7	5.7	1286	2 T33476	hypothetical prote
8	7	5.7	1301	1 A41622	protein-tyrosine-p
9	6	4.9	41	2 T12917	hypothetical prote
10	6	4.9	49	2 H81991	hypothetical prote
11	6	4.9	83	2 S37764	hypothetical prote
12	6	4.9	91	2 C33172	C-ORF-F protein -
13	6	4.9	94	2 D70245	hypothetical prote
14	6	4.9	98	2 G84296	hypothetical prote
15	6	4.9	116	2 G71168	hypothetical prote
16	6	4.9	118	2 S51973	hypothetical prote
17	6	4.9	124	2 I50095	MHC class II beta
18	6	4.9	129	2 S36999	ribosomal protein
19	6	4.9	144	2 B59304	methyl viologen-re
20	6	4.9	152	1 M6URIC	Spec1 protein - se
21	6	4.9	164	2 T00891	hypothetical prote
22	6	4.9	167	2 C64456	hypothetical prote
23	6	4.9	170	2 S76067	hypothetical prote
24	6	4.9	174	2 S73301	hypothetical prote
25	6	4.9	175	2 T14679	repressor of phase
26	6	4.9	185	2 D71854	fkbp-type peptidyl
27	6	4.9	185	2 C64660	peptidyl-prolyl ci
28	6	4.9	187	2 T14935	hypothetical prote
29	6	4.9	189	2 S44639	hypothetical prote

30 6 4.9 194 2 C72517 hypothetical prote
31 6 4.9 198 2 E75599 conserved hypothet
32 6 4.9 200 2 S28409 transsialidase-neu
33 6 4.9 219 2 G71343 hypothetical prote
34 6 4.9 220 2 S35075 class II histocomp
35 6 4.9 225 1 CYFGE epsilon-crystallin
36 6 4.9 226 2 G71873 hypothetical prote
37 6 4.9 231 2 I50467 MHC class II beta
38 6 4.9 240 2 T50628 hypothetical prote
39 6 4.9 243 1 TQ0021 ubiquinol--cytochr
40 6 4.9 244 2 D83020 hypothetical prote
41 6 4.9 245 2 S3867 DNA (cytosine)-met
42 6 4.9 246 2 I50094 MHC class II beta
43 6 4.9 247 2 I50093 MHC class II beta
44 6 4.9 247 2 I50465 MHC class II beta
45 6 4.9 247 2 I50466 MHC class II beta

ALIGNMENTS

RESULT 1

T08716
protein-tyrosine-phosphatase homolog DKFp566K0524.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
C:Accession: T08716
R:Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16472
A:Accession: T08716
A:Molecule type: mRNA
A:Residues: 1-398 <ANS>
A:Cross-references: EMBL:AL050040
A:Experimental source: fetal kidney; clone DKFp566K0524
C:Genetics:
A:Note: DKFp566K0524.1
C:Superfamily: protein-tyrosine-phosphatase homolog
F:161-379/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match 8.2%; Score 10; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 83 KWPDHGTPAS 92
Db 298 KWPDHGTPAS 307
|||||

RESULT 2

T22013
hypothetical protein F40D4.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T22013
R:Matthews, L.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19502
A:Accession: T22013
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-310 <WIL>
A:Cross-references: EMBL:Z81536; PTDN:CAB04364.1; GSPDB:GN00023; CESP:F40D4.5
A:Experimental source: clone F40D4
C:Genetics:
A:Gene: CESP:F40D4.5
A:Map position: 5
A:Introns: 45/1; 114/2; 194/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C13D9.6

Query Match 5.7%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ITGPLL 114
| | | | |
Db 234 ITGPLL 240

RESULT 3
A48711
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 2 - siime mold (Dictyostelium)
N:Alternate names: protein-tyrosine-phosphatase DdPTp2
C:Species: Dictyostelium discoideum
C:Date: 02-Jun-1995 #sequence_revision 08-Mar-1996 #text_change 11-Jun-1999
C:Accession: A48711; A53774
R:Ramalingam, R.; Shaw, D.R.; Ennis, H.L.
J. Biol. Chem. 268, 22680-22685, 1993
A:Title: Cloning and functional expression of a Dictyostelium discoideum protein tyrosine
A:Reference number: A48711; MUID:94043028
A:Accession: A48711
A:Molecule type: mRNA
A:Residues: 1-377 <RAM>
A:Cross-references: GB:L15420; NID:g290036; PIDN:AAA33242.1; PID:g290037
R:Howard, P.K.; Gamber, M.; Hunter, T.; Firtel, R.A.
Mol. Cell. Biol. 14, 5154-5164, 1994
A:Title: Regulation by protein-tyrosine phosphatase PTP2 is distinct from that by PTP1
A:Reference number: A53774; MUID:94309635
A:Accession: A53774
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA; DNA
A:Residues: 1-377 <HOW>
A:Cross-references: GB:L15420; NID:g290036; PIDN:AAA33242.1; PID:g290037
C:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: Dictyostelium protein-tyrosine-phosphatase, nonreceptor type 2; protein-
F:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:114-336/Domain: protein-tyrosine-phosphatase homology #status atypical <PTP>
F:281/Active site: Cys (phosphocysteine intermediate) #status predicted
F:287/Binding site: substrate phosphate (Arg) #status predicted

Query Match 5.7%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGTP 90
| | | | |
Db 249 WPDHGTP 255

RESULT 4
T21154
hypothetical protein F20E11.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T21154
R:Ainscough, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19383
A:Accession: T21154
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-397 <WIL>
A:Cross-references: EMBL:281508; PIDN:CAB04140.1; GSPDB:GN00023; CESP:F20E11.2
A:Experimental source: clone F20E11
C:Genetics:
A:Gene: CESP:F20E11.2
A:Map position: 5
A:Introns: 45/1; 114/2; 194/1; 281/1; 316/3
C:Superfamily: Caenorhabditis elegans hypothetical protein C13D9.6

Query Match 5.7%; Score 7; DB 2; Length 397;
Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ITGPLL 114
| | | | |
Db 234 ITGPLL 240

RESULT 5
G71568
probable (pet112) glu-tRNA gln amidotransferase (b chain) - Chlamydia trachomatis (se
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: G71568
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: G71568
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <ARN>
A:Cross-references: GB:AE001275; GB:AE001273; NID:g3328388; PIDN:AAC67594.1; PID:g3332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: gatB
C:Superfamily: PET112 protein

Query Match 5.7%; Score 7; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFQITQY 57
| | | | |
Db 94 NFQITQY 100

RESULT 6
B81722
glutamyl-tRNA(Gln) amidotransferase chain B TC0272 [Imported] - Chlamydia muridarum (C
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: B81722
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Ricke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: B81722
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <TET>
A:Cross-references: GB:AE002294; GB:AE002160; NID:g7190305; PIDN:AAF39140.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0272
C:Superfamily: PET112 protein

Query Match 5.7%; Score 7; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFQITQY 57
| | | | |
Db 94 NFQITQY 100

RESULT 7
T33476
hypothetical protein T27C10.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33476
 R:Zhu, H.J.; Graves, T.; Hawkins, M.
 Submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of *C. elegans* cosmid T27C10.
 A:Reference number: 221354
 A:Accession: T33476
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1286 <ZHU>
 A:Cross-references: EMBL:AF098504; PIDN:AAC67414.1; GSPDB:GN00019; CESP:T27C10.6
 A:Experimental source: strain Bristol N2; clone T27C10
 C:Genetics:
 A:Gene: CESP:T27C10.6
 A:Map position: 1
 A:Introns: 26/3; 79/1; 145/1; 182/2; 203/2; 269/1; 341/2; 397/3; 437/2; 719/1; 835/3

Query Match 5.7%; Score 7; DB 2; Length 1286;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 SLKEPLE 41
 |||||
 Db 998 SLKEPLE 1004

RESULT 8
 A1622
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 99A precursor - fruit fly (*Drosophila melanogaster*)
 N:Alternate names: phosphotyrosine phosphatase 99A
 C:Species: *Drosophila melanogaster*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A41622; A41214; B41215
 R:Harharan, I.K.; Chuang, P.T.; Rubin, G.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 11266-11270, 1991
 A:Title: Cloning and characterization of a receptor-class phosphotyrosine phosphatase gene
 A:Reference number: A41622; MUID:92107930
 A:Accession: A41622
 A:Molecule type: mRNA
 A:Residues: 1-1301 <HAR>
 A:Cross-references: GB:M81795; NID:g157293; PIDN:AAA28483.1; PID:g157294
 R:Yang, X.; Seow, K.T.; Bahri, S.M.; Oon, S.H.; Chia, W.
 Cell 67, 661-673, 1991
 A:Title: Two *Drosophila* receptor-like tyrosine phosphatase genes are expressed in a subset of the same cells
 A:Reference number: A41214; MUID:92034988
 A:Accession: A41214
 A:Molecule type: mRNA
 A:Residues: 1-585, 'R', 587-1049, 1120-1204, 'H', 1206-1301 <YAN>
 A:Cross-references: GB:M80464; NID:g157299; PIDN:AAA28486.1; PID:g157300
 A:Accession: B41214
 A:Molecule type: mRNA
 A:Residues: 1-585, 'R', 587-1049, 1290-1301 <YA2>
 A:Cross-references: GB:M80464
 R:Tian, S.S.; Tsoulfas, P.; Zinn, K.
 Cell 67, 675-685, 1991
 A:Title: Three receptor-linked protein-tyrosine phosphatases are selectively expressed in different subsets of *Drosophila* cells
 A:Reference number: A41215; MUID:92034989
 A:Accession: B41215
 A:Molecule type: mRNA
 A:Residues: 1-585, 'R', 587-1049, 1120-1184, 'S', 1186-1301 <TIA>
 A:Cross-references: GB:M80539
 C:Genetics:
 A:Gene: FlyBase:Ptp99A
 A:Cross-references: FlyBase:FBgn0004369
 C:Superfamily: protein-tyrosine-phosphatase, receptor type 99A; fibronectin type III repeat domain
 C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; transmembrane protein
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-1301/product: protein-tyrosine-phosphatase, receptor type 99A #status predicted <MA>
 F:30-1049, 1120-1301/product: protein-tyrosine-phosphatase, receptor type 99A, medium splice variant
 F:30-1049, 1290-1301/product: protein-tyrosine-phosphatase, receptor type 99A, short splice variant
 F:303-416/Domain: transmembrane #status predicted <TM>
 F:502-730/Domain: protein-tyrosine-phosphatase homology <PTP>
 F:789-1005/Domain: protein-tyrosine-phosphatase homology <PTP2>

F:1059-1091/Region: glutamine-rich
 F:682/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:688/Binding site: substrate phosphate (Arg) #status predicted
 Query Match 5.7%; Score 7; DB 1; Length 1301;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 84 WPDHCTP 90
 |||||
 Db 648 WPDHCTP 654

RESULT 9
 T12917
 hypothetical protein yosF - *Bacillus subtilis* phage SPBc2
 C:Species: *Bacillus subtilis* phage SPBc2
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
 C:Accession: T12917; H69925
 R:Lazarévic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
 submitted to the EMBL Data Library, August 1997
 A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBc2 protein
 A:Reference number: 217583
 A:Accession: T12917
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-41 <LAZ>
 A:Cross-references: EMBL:AF020713; NID:93025478; PID:g3025631; PIDN:AAC13126.1
 R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Broph, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: H69925
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-41 <KUN>
 A:Cross-references: GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13906.1; PID:ell1854
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yosF

Query Match 4.9%; Score 6; DB 2; Length 41;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 KKSTCK 72
 |||||
 Db 23 KKSTCK 18

RESULT 10
 H81991
 hypothetical protein NMA0012 [imported] - *Neisseria meningitidis* (strain 22491 serogr
 C:Species: *Neisseria meningitidis*
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: H81991
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
 Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
 Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: H81991
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-49 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83332.1; PID:g737879
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0012

Query Match 4.9%; Score 6; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 TAGVGR 122
DB 15 TAGVGR 20
|||||

RESULT 11
S37764
hypothetical protein - fruit fly (*Drosophila miranda*) transposon TRIM
C:Species: *Drosophila miranda*
C:Date: 19-May-1994 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: S37764
R:Steinmann, M.; Steinemann, S.
Chromosoma 101, 169-179, 1991
A:Title: Preferential X chromosomal location of TRIM, a novel transposable element of *D.*
A:Reference number: S37763; MUID:92164374
A:Accession: S37764
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-83 <STE>
A:Cross-references: EMBL:X59239; NID:g8718; PIDN:CAA11924.1; PID:g1335690
C:Genetics:
A:Gene: FlyBase:Dmir/TRIM
A:Cross-references: FlyBase:FBgn0004642
A:Mobile element: transposon TRIM

Query Match 4.9%; Score 6; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 REIEGG 24
DB 36 REIEGG 41
|||||

RESULT 12
C33172
C-ORF-F protein - vaccinia virus (strain Copenhagen)
N:Alternate names: B-ORF-G protein
C:Species: vaccinia virus
A:Note: host *Homo sapiens* (man)
C:Date: 19-Apr-1991 #sequence_revision 19-Apr-1991 #text_change 05-Jan-1996
C:Accession: C33172; H42529
R:Johnson, G.P.
submitted to GenBank, June 1990
A:Reference number: A33172
A:Accession: C33172
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-91 <JOH1>
A:Accession: H42529
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-91 <JOH2>
C:Genetics:
A:Note: the coding region for this protein is repeated in the viral genome

Query Match 4.9%; Score 6; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 KYRVYV 103
DB 31 KYRVYV 36
|||||

RESULT 13
D70245
hypothetical protein BBJ02 - Lyme disease spirochete plasmid J/lp38
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: D70245
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
A:Reference number: A70100; MUID:98065943
A:Accession: D70245
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-94 <KLE>
A:Cross-references: GB:AE000787; NID:g2690175; PIDN:AAC66115.1; PID:g2690207; TIGR:BB
A:Experimental source: strain B31
C:Genetics:
A:Gene: plasmid

Query Match 4.9%; Score 6; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 QIVKKS 69
DB 73 QIVKKS 78
|||||

RESULT 14
G84296
hypothetical protein Vng1425h [imported] - *Halobacterium* sp. NRC-1
C:Species: *Halobacterium* sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84296
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: G84296
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-98 <STO>
A:Cross-references: GB:AE004437; NID:g10580927; PIDN:AAG19739.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1425H

Query Match 4.9%; Score 6; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 PASVDF 95
DB 49 PASVDF 54
|||||

RESULT 15

G71168
Hypothetical protein PH0547 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: G71168
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137
A:Accession: G71168
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-116 <KAW>
A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29636.1; PID:g3256953
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0547
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0547

Query Match 4.9%; Score 6; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 IAMITR 19
|||||
Db 39 IAMITR 44

Search completed: March 30, 2002, 08:27:17
Job time: 559 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:27:19 ; Search time 10.3 Seconds
(without alignments)
434.282 Million cell updates/sec

Title: US-09-095-478A-8
Perfect score: 122
Sequence: 1 DFWMWNNCNCVIAITRE.....VRKSHITGLLVHCTAGVGR 122

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 100059 seqs, 36664827 residues

Word size : 0
Total number of hits satisfying chosen parameters: 100059
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries
Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	5.7	112	1 PT07_STYPL	P28199 styela plic
2	7	5.7	377	1 GATB_DICDI	P34138 dictyosteli
3	7	5.7	488	1 GATB_CHLMU	O9136 chlamydia m
4	7	5.7	488	1 GATB_CHLTR	O84007 chlamydia t
5	7	5.7	1301	1 PTP9_DROME	P35832 drosophila
6	7	5.7	2472	1 NCR2_MOUSE	Q9wu42 mus musculu
7	6	4.9	91	1 YVBG_VACCC	P20547 vaccinia vi
8	6	4.9	114	1 PT27_STYPL	P28219 styela plic
9	6	4.9	117	1 PT25_STYPL	P28217 styela plic
10	6	4.9	118	1 YAE6_YEAST	P39724 saccharomyc
11	6	4.9	129	1 RL31_SULAC	P38618 sulfolobus
12	6	4.9	144	1 CSF2_FEICA	O62757 felis silve
13	6	4.9	152	1 SP1A_STRPU	P04109 strongyloce
14	6	4.9	174	1 YC21_PORPU	P51380 porphyra pu
15	6	4.9	185	1 SLVD_HELPJ	O92k89 helicobacte
16	6	4.9	185	1 SLVD_HELPY	O25748 helicobacte
17	6	4.9	189	1 YPT2_CAEEL	P41880 caenorhabdi
18	6	4.9	219	1 Y284_TREPA	O83308 treponema p
19	6	4.9	243	1 CY1_EUGGR	P20114 euclena gra
20	6	4.9	251	1 HIS6_SULSO	O33774 sulfolobus
21	6	4.9	259	1 UPKB_BOVIN	P38573 bos taurus
22	6	4.9	259	1 UPKB_MUSVI	P30413 mustela vis
23	6	4.9	279	1 ATPC_MYCGE	P47640 mycoplasma
24	6	4.9	306	1 YEDA_ECOLI	P09185 escherichia
25	6	4.9	309	1 Y209_MYCPN	P75485 mycoplasma
26	6	4.9	323	1 CRO_RANCA	P17264 rana catesb
27	6	4.9	323	1 CRO_RANTE	P02532 rana tempor
28	6	4.9	327	1 AR71_RAT	P38918 rattus norv
29	6	4.9	330	1 AR72_HUMAN	O43488 homo sapien
30	6	4.9	339	1 REP_BACAM	P13963 bacillus am
31	6	4.9	371	1 Y028_ARCFU	O30207 archaeoglob
32	6	4.9	379	1 METX_MYCTU	O53391 mycobacteri
33	6	4.9	382	1 METX_MYCLE	O32874 mycobacteri

34 TAL_SYNY3 391 1 TAL_SYNY3 P72797 synechocyst
35 RRP_NDVA 395 1 RRP_NDVA P16073 newcastle d
36 YM65_MYCTU 409 1 YM65_MYCTU Q50697 mycobacteri
37 RFL_METJA 421 1 RFL_METJA Q58239 methanococc
38 MYCS_RAT 429 1 MYCS_RAT P23999 rattus norv
39 RHGA_ASPAC 440 1 RHGA_ASPAC O00001 aspergillus
40 GSPB_BACST 445 1 GSPB_BACST P13376 bacillus st
41 CS66_WHEAT 469 1 CS66_WHEAT P45226 triticum ae
42 ATPB_MYCGA 471 1 ATPB_MYCGA P33253 mycoplasma
43 MM10_MOUSE 476 1 MM10_MOUSE O55123 mus musculu
44 MM03_MOUSE 477 1 MM03_MOUSE P28862 mus musculu
45 DHAB_RHINE 487 1 DHAB_RHINE P54222 rhizobium m

ALIGNMENTS

RESULT 1
PT07_STYPL
ID PT07_STYPL STANDARD; PRT; 112 AA.
AC P28199;
DT 01-DEC-1992 (Rel. 24, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE 7 (EC 3.1.3.48) (FRAGMENT).
GN STY 7.
OS Styela plicata (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Styelidae; Styela.
OX NCBI_TaxID=7726;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91139172; Pubmed=1704870;
PA Matthews R.J., Flores E., Thomas M.L.;
RT "Protein tyrosine phosphatase domains from the protochordate Styela
plicata".
RL Immunogenetics 33:33-41(1991).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Matthews R.J., Flores E., Thomas M.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBAJ databases.
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37992; AAA29825.1; --
CC HSSP; P28827; IRPM.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phphatase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00012; PTPC_DSPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; PARTIAL.
DR PROSITE; PS00356; TYR_PHOSPHATASE_2; PARTIAL.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
FT HYDROLASE.
FT NON_TER 1 112
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 13299 MW; 6F19D514C685D9FC CRC64;

Query Match 5.7%; Score 7; DB 1; Length 112;
Best local Similarity 100.0%; Pred. No. 1.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 84 WPDHGTP 90
DB 80 WPDHGTP 86

RESULT 2
PTP2_DICDI
ID PTP2_DICDI STANDARD; PRT; 377 AA.
AC P34138;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE 2 (EC 3.1.3.48) (PROTEIN-TYROSINE-
DE PHOSPHATE PHOSPHOHYDROLASE 2) (PTPA).
GN PTPB OR PTP2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=46689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=94043028; PubMed=8226777;
RA Ramalingam R., Shaw D.R., Ennis H.;
RT "Cloning and functional expression of a Dictyostelium discoideum
RT protein tyrosine phosphatase";
RL J. Biol. Chem. 268:22680-22685(1993).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L15420; AAA33242.1; -
DR HSP; P18031; 2HNP.
DR DICTYDB; D005054; ptpb.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF000242; TYR_prot_phptase.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
FT DOMAIN 78 86 POLY-ASP.
FT DOMAIN 95 103 POLY-ASN.
FT ACT_SITE 281 281 BY SIMILARITY.
SQ SEQUENCE 377 AA; 43488 MW; 5B08F6EB54829FD9 CRC64;

Query Match 5.7%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGTP 90
DB 249 WPDHGTP 255

RESULT 3
GATB_CHLMU
ID GATB_CHLMU STANDARD; PRT; 488 AA.
AC Q9PLJ6;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

QY 84 WPDHGTP 90
DB 249 WPDHGTP 255

Query Match 5.7%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGTP 90
DB 249 WPDHGTP 255

Query Match 5.7%; Score 7; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57
DB 94 NFOITQY 100

RESULT 4
GATB_CHLTR
ID GATB_CHLTR STANDARD; PRT; 488 AA.
AC O84007;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR CT004.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,

```

```

DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR CT0272.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSMIDATION OF MISACYLATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GATB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF002294; AAF39140.1; -
DR TIGR; TC0272; -
DR InterPro; IPR001773; Gln_amidotransf_B.
DR Pfam; PF01162; PET112; 1.
DR PROSITE; PS01234; GATB; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 488 AA; 54759 MW; B1966520739D041E CRC64;

Query Match 5.7%; Score 7; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57
DB 94 NFOITQY 100

RESULT 4
GATB_CHLTR
ID GATB_CHLTR STANDARD; PRT; 488 AA.
AC O84007;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR CT004.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,

```


RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans:

RL Chlamydia trachomatis.";

CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED

CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-

CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE

CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH

CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).

CC + CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP

CC + PHOSPHATE + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE.

CC -!- SUBUNIT: HETERODIMER OF A, B AND C SUBUNITS (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE GATB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: AE001275; AAC67594.1; -

CC InterPro: IPR001773; Gln_amidotransf_B.

CC Pfam: PF01162; PET112.1

CC PROSITE: PS01234; GATB.1.

KW Protein biosynthesis; Ligase; Complete proteome.

SO SEQUENCE 488 AA; 55018 MW; A6F9D586042DA112 CRC64;

Query Match 5.78; Score 7; DB 1; Length 488;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 51 NFIQTQY 57

Db 94 NFIQTQY 100

|||||

RESULT 5

PTP9_DROME STANDARD; PRT; 1301 AA.

AC P35832;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROTEIN-TYROSINE PHOSPHATASE 99A PRECURSOR (EC 3.1.3.48) (RECEPTOR-

DE LINKED PROTEIN-TYROSINE PHOSPHATASE 99A).

GN PTP99A.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

PP1 SEQUENCE FROM N.A.

RC TISSUE-Eye imaginal disk;

RX MEDLINE=92107930; PubMed=16623390;

RA Hariharan I.K., Chuang P.-T., Rubin G.M.;

RT "Cloning and characterization of a receptor-class phosphotyrosine

RT phosphatase gene expressed on central nervous system axons in

RT Drosophila melanogaster.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:11266-11270(1991).

RP [2]

RP SEQUENCE FROM N.A.

RC TISSUE-Embryo;

RX MEDLINE=92034989; PubMed=1657402;

RA Tian S.-S., Tsoulfas P., Zinn K.;

RT "Three receptor-linked protein-tyrosine phosphatases are selectively

RT expressed on central nervous system axons in the Drosophila embryo.";

RL Cell 67:675-685(1991).

RP [3]

RP SEQUENCE FROM N.A.

TISSUE-Embryo;

RC MEDLINE=92034988; PubMed=1657401;

RA Wang X., Seow K.T., Bahri S.M., Oon S.H., Chia W.;

RT "Two Drosophila receptor-like tyrosine phosphatase genes are

RT expressed in a subset of developing axons and pioneer neurons in the

RT embryonic CNS.";

RL Cell 67:661-673(1991).

CC -!- FUNCTION: MAY PLAY A KEY ROLE IN SIGNAL TRANSDUCTION AND GROWTH

CC -!- CONTROL.

CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =

CC PROTEIN TYROSINE + ORTHOPHOSPHATE.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- ALTERNATIVE PRODUCTS: TWO ISOMERS THAT DIFFER IN THEIR C-TERMINAL

CC TAILS ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND

CC PIONEER NEURONS IN THE EMBRYO.

CC -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS

CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: M81795; AAA28483.1; -

CC EMBL: M80539; AAA28485.1; -

CC EMBL: M80464; AAA28486.1; -

CC PIR: A41622; A41622.

CC PIR: A41214; A41214.

CC PIR: B41214; B41214.

CC PIR: B41215; B41215.

CC HSSP: P18052; 1VFO.

CC FlyBase: FBgn0004369; Ptp99A.

CC InterPro: IPR001777; FN_III.

CC InterPro: IPR000387; Tyr_phosphatase.

CC Pfam: PF00041; fn3; 2.

CC Pfam: PF0102; Y_phosphatase; 2.

CC PRINTS: PR00014; FNTYPRIII

CC PRINTS: PR00700; PTPYPHPTASE.

CC SMART: SM00060; FN3; 2.

CC SMART: SM00194; PTPC; 2.

CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.

CC PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.

CC PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.

KW Transmembrane; Hydrolase; Duplication; Signal; Alternative splicing;

KW Repeat.

FT SIGNAL 1 29

FT CHAIN 30 1301

FT DOMAIN 30 392

FT TRANSMEM 393 416

FT DOMAIN 417 1301

FT DOMAIN 64 168

FT DOMAIN 169 268

FT DOMAIN 269 368

FT DOMAIN 497 747

FT DOMAIN 748 975

FT ACT_SITE 682 682

FT DOMAIN 1076 1091

FT CARBOHYD 33 33

FT CARBOHYD 176 176

FT CARBOHYD 212 212

FT CARBOHYD 278 278

FT CARBOHYD 322 322

FT CARBOHYD 336 336

FT VARSPIC 1050 1119

FT CONFLICT 586 586

FT CONFLICT 1205 1205

SO SEQUENCE I301 AA; 145336 MW; 8241E3E19A4CA5BD CRC64;

POTENTIAL.

PROTEIN-TYROSINE PHOSPHATASE 99A.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOPLASMIC (POTENTIAL).

FIBRONECTIN TYPE-III 1.

FIBRONECTIN TYPE-III 2.

FIBRONECTIN TYPE-III 3.

PROTEIN-TYROSINE PHOSPHATASE 1.

PROTEIN-TYROSINE PHOSPHATASE 2.

BY SIMILARITY.

POLY-GLN.

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

MISSING (IN SHORT ISOFORM).

P -> R (IN REF. 2 AND 3).

N -> H (IN REF. 3).

```
Query Match 5.7%; Score 7; DB 1; Length 1301;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGT 90
Db 648 WPDHGT 654

RESULT 6
ID NCR2_MOUSE STANDARD; PRT; 2472 AA.
AC Q9WU42; Q9WU43; Q9WU41;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF
DE RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (THYROID-,
DE RETINOIC ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-
DE ASSOCIATING FACTOR) (TRAC).
GN NCOR2 OR SMRT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA AND ISOFORM BETA).
RC TISSUE=Spleen, and Brain;
RA MEDLINE=99178941; PubMed=10077563;
RT Orthelich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Embryo;
RA MEDLINE=99199215; PubMed=10097068;
RT Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;
RT "SMRT, a silencing mediator for retinoid and thyroid hormone
RT receptors-extended isoform that is more related to the nuclear
RT receptor corepressor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIb.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY
CC EMBRYOS.
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC DOMAINS (ID1 AND ID2).
CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES
CC FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC SPECIFICITY.
CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).
CC -!- SIMILARITY: CONTAINS 2 CORN BOX.
CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; AF113001; AAD20944.1; -
DR DR EMBL; AF113002; AAD20945.1; -
DR DR EMBL; AF125671; AAD22972.1; -
DR DR MGD; MGI:1337080; Ncor2.
DR DR InterPro; IPR001005; Myb_DNA_bind.
DR DR Pfam; PF00249; myb_DNA-binding; 2.
DR DR SMART; SM00395; SANT; 2.
DR DR PROSITE; PS50090; MYB_3; 1.
KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW Coiled coil; Alternative splicing.
FT DOMAIN 165 207 COILED COIL (POTENTIAL).
FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT DNA_BIND 429 474 SANT-A (POTENTIAL).
FT DNA_BIND 608 653 SANT-B (POTENTIAL).
FT DOMAIN 492 560 COILED COIL (POTENTIAL).
FT DOMAIN 652 682 COILED COIL (POTENTIAL).
FT DOMAIN 775 804 PRO-RICH.
FT DOMAIN 989 999 PRO-RICH.
FT DOMAIN 1351 1357 PRO-RICH.
FT DOMAIN 2094 2098 CORNR BOX OF ID1.
FT DOMAIN 2296 2300 CORNR BOX OF ID2.
FT DOMAIN 494 507 POLY-GLN.
FT DOMAIN 1615 1619 POLY-ALA.
FT DOMAIN 2434 2437 POLY-PRO.
FT VARSPLIC 36 254 MISSING (IN ISOFORM BETA).
FT CONFLICT 176 176 M -> RL (IN REF. 2).
FT CONFLICT 396 402 PPMYDA -> RHVRR (IN REF. 2).
FT CONFLICT 555 555 D -> H (IN REF. 1; AAD20944).
FT CONFLICT 756 756 T -> M (IN REF. 1; AAD20944).
FT CONFLICT 785 785 V -> A (IN REF. 2).
FT CONFLICT 806 846 HILPHRLTLTRNKKPRLLQLPRORMPRSRSLRPRSMVE
-> PSPAAPATVVDKDEQAPAPAPQTEDAKQKSEAEI
DVG (IN REF. 2).
FT CONFLICT 856 856 E -> K (IN REF. 1; AAD20945).
FT CONFLICT 859 859 E -> K (IN REF. 1; AAD20945).
FT CONFLICT 867 867 E -> K (IN REF. 1; AAD20945).
FT CONFLICT 895 895 E -> K (IN REF. 1; AAD20945).
FT CONFLICT 916 916 S -> F (IN REF. 1; AAD20944).
FT CONFLICT 975 975 I -> IQ (IN REF. 1; AAD20944).
FT CONFLICT 1046 1063 PKLTPPPRWSSGLPPFI -> QSYRLSPHAGRLPSH
(IN REF. 2).
FT CONFLICT 1073 1080 PHADESA -> TRADPL (IN REF. 2).
FT CONFLICT 1133 1133 MISSING (IN REF. 2).
FT CONFLICT 1149 1149 MISSING (IN REF. 2).
FT CONFLICT 1157 1157 G -> E (IN REF. 2).
FT CONFLICT 1172 1201 GSATSGTKGLSTRAADGPGSYRGSIHG -> APPPVEA
SPRASQYGCRRPQLQRLYHPR (IN REF. 2).
FT CONFLICT 1696 1696 A -> S (IN REF. 2).
FT CONFLICT 1855 1857 MISSING (IN REF. 2).
FT CONFLICT 1909 1909 A -> P (IN REF. 2).
FT CONFLICT 1913 1913 A -> G (IN REF. 2).
FT CONFLICT 1923 1923 G -> A (IN REF. 2).
FT CONFLICT 1956 1956 N -> S (IN REF. 2).
FT CONFLICT 1968 1968 A -> G (IN REF. 2).
FT CONFLICT 2195 2196 TA -> AV (IN REF. 2).
FT CONFLICT 2213 2214 LE -> SK (IN REF. 2).
FT CONFLICT 2224 2224 T -> A (IN REF. 2).
SQ SEQUENCE 2472 AA; 270856 MW; 2A58F4DF7B79285B CRC64;

Query Match 5.7%; Score 7; DB 1; Length 2472;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 PDHGTGA 91
Db 2143 PDHGTGA 2149
```

```

CC CC RESULT 7
CC CC YVGB_VACCC
CC CC ID YVGB_VACCC STANDARD; PRT; 91 AA.
CC CC P20547;
CC CC 01-FEB-1991 (Rel. 17, Created)
CC CC 01-FEB-1991 (Rel. 17, Last sequence update)
CC CC 20-AUG-2001 (Rel. 40, Last annotation update)
CC CC HYPOTHETICAL 11.2 KDA PROTEIN.
CC CC B ORF G AND C ORF F.
CC CC OS Vaccinia virus (Strain Copenhagen).
CC CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
CC CC Orthopoxvirus.
CC CC NCBI_TaxID=10249;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=91021027; PubMed=2219772;
CC CC Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
CC CC Paoletti E.;
CC CC "The complete DNA sequence of vaccinia virus.";
CC CC Virology 179:247-266(1990).
CC CC [2]
CC CC COMPLETE GENOME.
CC CC Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
CC CC Paoletti E.;
CC CC "Appendix to 'The complete DNA sequence of vaccinia virus'.";
CC CC Virology 179:517-563(1990).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M35027; AAA48224.1; -
CC CC EMBL; M35027; AAA47977.1; -
CC CC PIR; C33172; C33172.
CC CC PIR; H42529; H42529.
CC CC Hypothetical protein.
CC CC SEQUENCE 91 AA; 11195 MW; 3621A57A685C9421 CRC64;
CC CC -----
CC CC Query Match 4.9%; Score 6; DB 1; Length 91;
CC CC Best Local Similarity 100.0%; Pred. No. 17;
CC CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC -----
CC CC QY 98 KVRVY 103
CC CC 111111
CC CC DB 31 KVRVY 36
CC CC -----
CC CC RESULT 8
CC CC PT27_STYPL
CC CC ID PT27_STYPL STANDARD; PRT; 114 AA.
CC CC AC P28219;
CC CC 01-DEC-1992 (Rel. 24, Created)
CC CC 01-DEC-1992 (Rel. 24, Last sequence update)
CC CC 01-DEC-1992 (Rel. 24, Last annotation update)
CC CC PROTEIN-TYROSINE PHOSPHATASE 27 (EC 3.1.3.48) (FRAGMENT).
CC CC STY 27.
CC CC Styela plicata (Sea squirt).
CC CC OS Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
CC CC OC Stolidobranchia; Styelidae; Styela.
CC CC NCBI_TaxID=7726;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=91139172; PubMed=1704870;
CC CC Matthews R.J., Flores E., Thomas M.L.;
CC CC "Protein tyrosine phosphatase domains from the protochordate Styela
CC CC plicata.";
CC CC Immunogenetics 33:33-41(1991).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M38010; AAA29843.1; -
CC CC HSSP; P18052; 1YFO.
CC CC InterPro; IPR000340; DS_phosphatase.
CC CC InterPro; IPR000387; TYR_phosphatase.
CC CC Pfam; PF000102; Y_phosphatase; 1.
CC CC SMART; PS00012; PTPC_DSpc; 1.
CC CC -----
CC CC RESULT 9
CC CC PT25_STYPL
CC CC ID PT25_STYPL STANDARD; PRT; 117 AA.
CC CC AC P28217;
CC CC 01-DEC-1992 (Rel. 24, Created)
CC CC 01-DEC-1992 (Rel. 24, Last sequence update)
CC CC 01-DEC-1992 (Rel. 24, Last annotation update)
CC CC PROTEIN-TYROSINE PHOSPHATASE 25 (EC 3.1.3.48) (FRAGMENT).
CC CC STY 25.
CC CC Styela plicata (Sea squirt).
CC CC OS Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
CC CC OC Stolidobranchia; Styelidae; Styela.
CC CC NCBI_TaxID=7726;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=91139172; PubMed=1704870;
CC CC Matthews R.J., Flores E., Thomas M.L.;
CC CC "Protein tyrosine phosphatase domains from the protochordate Styela
CC CC plicata.";
CC CC Immunogenetics 33:33-41(1991).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M38012; AAA29819.1; -
CC CC HSSP; P28827; 1RPM.
CC CC InterPro; IPR000340; DS_phosphatase.
CC CC InterPro; IPR000387; TYR_phosphatase.
CC CC Pfam; PF000242; TYR_prot_phphatase.
CC CC SMART; SM00012; Y_phosphatase; 1.
CC CC SMART; PS00012; PTPC_DSpc; 1.
CC CC PROSITE; PS00383; TYR_PHOSPHATASE_1; PARTIAL.
CC CC PROSITE; PS00556; TYR_PHOSPHATASE_2; PARTIAL.
CC CC PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
CC CC KW Hydrolase.
CC CC FT NON_TER 1
CC CC FT NON_TER 114 114
CC CC SQ SEQUENCE 114 AA; 13375 MW; 80D335FA4C72CEAF CRC64;
CC CC -----
CC CC Query Match 4.9%; Score 6; DB 1; Length 114;
CC CC Best Local Similarity 100.0%; Pred. No. 21;
CC CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC -----
CC CC QY 83 KWPDHG 88
CC CC 111111
CC CC DB 79 KWPDHG 84
CC CC -----

```

```

CC CC RESULT 7
CC CC YVGB_VACCC
CC CC ID YVGB_VACCC STANDARD; PRT; 91 AA.
CC CC P20547;
CC CC 01-FEB-1991 (Rel. 17, Created)
CC CC 01-FEB-1991 (Rel. 17, Last sequence update)
CC CC 20-AUG-2001 (Rel. 40, Last annotation update)
CC CC HYPOTHETICAL 11.2 KDA PROTEIN.
CC CC B ORF G AND C ORF F.
CC CC OS Vaccinia virus (Strain Copenhagen).
CC CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
CC CC Orthopoxvirus.
CC CC NCBI_TaxID=10249;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=91021027; PubMed=2219722;
CC CC Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
CC CC Paoletti E.;
CC CC "The complete DNA sequence of vaccinia virus.";
CC CC Virology 179:247-266(1990).
CC CC [2]
CC CC COMPLETE GENOME.
CC CC Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
CC CC Paoletti E.;
CC CC "Appendix to 'The complete DNA sequence of vaccinia virus'.";
CC CC Virology 179:517-563(1990).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M35027; AAA48224.1; -
CC CC EMBL; M35027; AAA47977.1; -
CC CC PIR; C33172; C33172.
CC CC PIR; H42529; H42529.
CC CC Hypothetical protein.
CC CC SEQUENCE 91 AA; 11195 MW; 3621A57A685C9421 CRC64;
CC CC
CC CC Query Match 4.9%; Score 6; DB 1; Length 91;
CC CC Best Local Similarity 100.0%; Pred. No. 17;
CC CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 98 KVRVY 103
CC CC
CC CC DB 31 KVRVY 36
CC CC
CC CC
CC CC RESULT 8
CC CC PT27_STYPL
CC CC ID PT27_STYPL STANDARD; PRT; 114 AA.
CC CC AC P28219;
CC CC 01-DEC-1992 (Rel. 24, Created)
CC CC 01-DEC-1992 (Rel. 24, Last sequence update)
CC CC 01-DEC-1992 (Rel. 24, Last annotation update)
CC CC PROTEIN-TYROSINE PHOSPHATASE 27 (EC 3.1.3.48) (FRAGMENT).
CC CC STY 27.
CC CC Styela plicata (Sea squirt).
CC CC OS Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
CC CC OC Stolidobranchia; Styelidae; Styela.
CC CC NCBI_TaxID=7726;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=91139172; PubMed=1704870;
CC CC Matthews R.J., Flores E., Thomas M.L.;
CC CC "Protein tyrosine phosphatase domains from the protochordate Styela
CC CC plicata.";
CC CC Immunogenetics 33:33-41(1991).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M38010; AAA29843.1; -
CC CC HSSP; P18052; 1YFO.
CC CC InterPro; IPR000340; DS_phosphatase.
CC CC InterPro; IPR000387; TYR_phosphatase.
CC CC Pfam; PF000102; Y_phosphatase; 1.
CC CC SMART; SM00012; PTPC_DSPC; 1.
CC CC
CC CC RESULT 9
CC CC PT25_STYPL
CC CC ID PT25_STYPL STANDARD; PRT; 117 AA.
CC CC AC P28217;
CC CC 01-DEC-1992 (Rel. 24, Created)
CC CC 01-DEC-1992 (Rel. 24, Last sequence update)
CC CC 01-DEC-1992 (Rel. 24, Last annotation update)
CC CC PROTEIN-TYROSINE PHOSPHATASE 25 (EC 3.1.3.48) (FRAGMENT).
CC CC STY 25.
CC CC Styela plicata (Sea squirt).
CC CC OS Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
CC CC OC Stolidobranchia; Styelidae; Styela.
CC CC NCBI_TaxID=7726;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=91139172; PubMed=1704870;
CC CC Matthews R.J., Flores E., Thomas M.L.;
CC CC "Protein tyrosine phosphatase domains from the protochordate Styela
CC CC plicata.";
CC CC Immunogenetics 33:33-41(1991).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M38010; AAA29843.1; -
CC CC HSSP; P18052; 1YFO.
CC CC InterPro; IPR000340; DS_phosphatase.
CC CC InterPro; IPR000387; TYR_phosphatase.
CC CC Pfam; PF000102; Y_phosphatase; 1.
CC CC SMART; SM00012; PTPC_DSPC; 1.

```

```

DR PROSITE; PS00383; TYR_PHOSPHATASE_1; PARTIAL.
DR PROSITE; PS00356; TYR_PHOSPHATASE_2; PARTIAL.
DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
FT NON_TER 1 1
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 13422 MW; 9CAE0F5D70B7EF28 CRC64;

Query Match
Best Local Similarity 4.9%; Score 6; DB 1; Length 117;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KWPDPHG 88
DB 82 KWPDPHG 87

RESULT 10
YAE6_YEAST
ID YAE6_YEAST STANDARD; PRT; 118 AA.
AC P39724;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHEICAL 13.4 KDA PROTEIN IN ACS1-GCV3 INTERGENIC REGION.
GN YAL046C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U12980; AAC04985.1; -.
CC SGD; S0000044; YAL046C.
CC InterPro; IPR002634; BOLA.
CC Pfam; PF01722; BOLA; 1.
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 13356 MW; 244F5FF2052FF410 CRC64;

Query Match
Best Local Similarity 4.9%; Score 6; DB 1; Length 118;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KKSTGK 72
DB 108 KKSTGK 113

RESULT 11
RL31_SULAC
ID RL31_SULAC STANDARD; PRT; 129 AA.
AC P38618;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)

```

```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L31E.
GN RPL31E.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=95226466; PubMed=7711082;
RA Moll R., Schaefer G., Schmidtke S.;
RT "Nucleotide sequence of a gene cluster encoding ribosomal proteins in
the thermophilic crenarchaeon Sulfolobus acidocaldarius."
RL Bicchim. Biophys. Acta 1261:315-318(1995).
CC -!- SIMILARITY: BELONGS TO THE L31E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; X77509; CAAS4639.1; -.
CC F01R; S41964; S41964.
CC InterPro; IPR000054; Ribosomal_L31e.
CC ProDom; PD006030; Ribosomal_L31e; 1.
DR PROSITE; PS01144; RIBOSOMAL_L31E; FALSE_NEG.
KW Ribosomal protein.
SQ SEQUENCE 129 AA; 15196 MW; 56ACF37C68E076A0 CRC64;

Query Match
Best Local Similarity 4.9%; Score 6; DB 1; Length 129;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 IKVRY 102
DB 70 IKVRY 75

RESULT 12
CSF2_FELCA
ID CSF2_FELCA STANDARD; PRT; 144 AA.
AC O62757;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR PRECURSOR (GM-CSF)
HE (COLONY-STIMULATING FACTOR) (CSF).
GN CSF2.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RA Hughes K.J., O'Reilly K.L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CYTOKINE THAT STIMULATES THE GROWTH AND DIFFERENTIATION
CC OF HEMATOPOIETIC PRECURSOR CELLS FROM VARIOUS LINEAGES, INCLUDING
CC GRANULOCYTES, MACROPHAGES, EOSINOPHILS AND ERYTHROCYTES (BY
CC SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)

```

```
CC EMBL: AF053007; AAC06041.1; -
DR InterPro: IPR000773; GM_CSF.
DR Pfam: PF01109; GM_CSF; 1.
DR PRINTS: PR00693; GMCFACTOR.
DR PRODom: PD007349; GM_CSF; 1.
DR SMART: SM00040; CSF2; 1.
DR PROSITE: PS00702; GM_CSF; 1.
KW Cytokine; Growth factor; Glycoprotein; Signal.
FT SIGNAL 1 17
FT CHAIN 18 144
FT
FT DISULFID 71 113
FT BY SIMILARITY.
FT DISULFID 105 138
FT BY SIMILARITY.
FT CARBOHYD 44 44
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 45 45
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 144 AA; 16505 MW; 42C5BF5F0235DA55 CRC64;

Query Match 4.9%; Score 6; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SLKEPL 40
DB 89 SLKEPL 94

RESULT 13
SP1A_STRPU STANDARD; PRT; 152 AA.
AC P04109;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPC1 1A PROTEIN.
GN SPC1
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Gastrula;
RX MEDLINE=88011349; PubMed=3116272;
RA Hardin P.E., Klein W.H.;
RT "Unusual sequence conservation in the 5' and 3' untranslated regions
of the sea urchin spec mRNAs.";
RL J. Mol. Evol. 25:126-133(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86115267; PubMed=2935638;
RA Hardin S.H., Carpenter C.D., Hardin P.E., Bruskin A.M., Klein W.H.;
RT "Structure of the Speg1 gene encoding a major calcium-binding protein
in the embryonic ectoderm of the sea urchin, Strongylocentrotus
purpuratus.";
RL J. Mol. Biol. 186:243-255(1985).
RN [3]
RP SEQUENCE OF 67-152 FROM N.A.
RX MEDLINE=84130177; PubMed=6697391;
RA Carpenter C.D., Bruskin A.M., Hardin P.E., Keast M.J., Anstrom J.A.,
RA Tyner A.L., Brandhorst B.P., Klein W.H.;
RT "Novel proteins belonging to the troponin C superfamily are encoded
by a set of mRNAs in sea urchin embryos.";
RL Cell 36:663-671(1984).
RN [4]
RP FUNCTION: CALCIUM-BINDING PROTEIN INVOLVED IN LARVAL DEVELOPMENT
AND METAMORPHOSIS. LIKELY TO FUNCTION AS CALCIUM BUFFERS
MEDATING THE TRANSPORT OF CALCIUM FROM THE SEA WATER TO THE
BLASTOCOELE WHERE CALCIUM IS REQUIRED FOR SKELETON FORMATION.
CC -!- SIMILARITY: BELONGS TO THE YCF21 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC LATE STAGE EMBRYO AND LARVA.
CC DEVELOPMENTAL STAGE: ACCUMULATE IN EMBRYOS AND LARVAE, BUT NOT IN
CC ADULTS.
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M32446; AAA30073.1; -
DR EMBL: X03287; CAA27036.1; -
DR EMBL: X03288; CAA27036.1; JOINED.
DR EMBL: X03289; CAA27036.1; JOINED.
DR EMBL: X03290; CAA27036.1; JOINED.
DR EMBL: X03291; CAA27036.1; JOINED.
DR EMBL: X03292; CAA27036.1; JOINED.
DR EMBL: K01649; AAA30076.1; -
DR PIR: A03031; MCURIC.
DR HSSP: P02593; 1AK8.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 4.
DR SMART: SM00054; EFh; 4.
DR PROSITE: PS00018; EF_HAND; 4.
KW Calcium-binding; Repeat.
FT CA_BIND 23 34
FT CA_BIND 59 70
FT CA_BIND 97 108
FT CA_BIND 133 144
FT CA_BIND 132 132
FT CONFLICT A -> P (IN REF. 1).
SQ SEQUENCE 152 AA; 17466 MW; B7C68B9A68F261B2 CRC64;

Query Match 4.9%; Score 6; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 KSTGKS 73
DB 40 KSTGKS 45

RESULT 14
YC2L_PORPU STANDARD; PRT; 174 AA.
AC P51380;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 20.6 KDA PROTEIN YCF21 (ORF174).
GN YCF21
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AVONPORT;
RA Reith M.E., Munnolland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
RN [2]
RP SIMILARITY: BELONGS TO THE YCF21 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

CC or send an email to license@isb-sib.ch).

CC EMBL; U38804; AAC08266.1; -
DR Mendel; 10371; PORpu:ycf21.1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 174 AA; 20632 MW; 7FF9D65664994E0E CRC64;

Query Match 4.9%; Score 6; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LLENFQ 53
| | | | |
Db 169 LLENFQ 174

RESULT 15

SLYD_HELPJ STANDARD; PRT; 185 AA.
AC Q9ZK89;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE SLYD (EC 5.2.1.8)
DE (PPIASE) (ROTAMASE).
GN SLYD OR JHP1052.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001533; AAD06631.1; -
DR InterPro; IPR001179; FKBP_PPIase.
DR Pfam; PF00254; FKBP; 1.
DR PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR PROSITE; PS00454; FKBP_PPIASE_2; FALSE_NEG.
DR PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW Isomerase; Rotamase; Complete proteome.
FT DOMAIN 1 99 PPIASE, FKBP-TYPE.
FT DOMAIN 158 185 GLY-RICH.
FT SEQUENCE 185 AA; 20100 MW; 2D48E2AC450B0CB CRC64;

Query Match 4.9%; Score 6; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 KEPLF 42
| | | | |
Db 34 KEPLF 39

Search completed: March 30, 2002, 08:31:33
Job time: 254 sec.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:26:48 ; Search time 27.01 Seconds
(without alignments)
660.690 Million cell updates/sec

Title: US-09-095-478A-8
Perfect score: 122
Sequence: 1 DWGMMWNNCNCVIAITRE.....VRKSHITGPLLVCAGVGR 122

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : SPTREMBL17.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	18.9	426	11	O55082 mus musculus
2	10	8.2	398	4	Q9Y406 homo sapien
3	7	5.7	52	7	Q9TPH8 xiphophorus
4	7	5.7	116	11	Q9D792 mus musculus
5	7	5.7	158	12	Q73103 human immun
6	7	5.7	244	7	O62884 xiphophorus
7	7	5.7	310	5	Q9XV30 caenorhabdi
8	7	5.7	351	10	Q9SST9 scaparia du
9	7	5.7	378	10	Q9M7Q0 porphyra pu
10	7	5.7	378	10	Q9M7P9 porphyra pu
11	7	5.7	378	10	Q9M7P8 porphyra pu
12	7	5.7	397	5	Q9XV85 caenorhabdi
13	7	5.7	407	2	Q9AA59 caulobacter
14	7	5.7	446	5	Q9YIX4 ephydalia f
15	7	5.7	488	5	Q9NL12 branchiosto
16	7	5.7	542	5	Q24198 drosophila
17	7	5.7	542	5	Q9VIB9 drosophila
18	7	5.7	991	2	Q9L773 xanthomonas
19	7	5.7	1214	5	Q9VAL3 drosophila

7	5.7	1286	5	Q9TZM3	Q9tzm3 caenorhabdi
6	4.9	41	2	O31883	O31883 bacillus su
6	4.9	41	9	O64165	O64165 bacterioph
22	4.9	48	2	Q48270	Q48270 helicobacte
23	4.9	49	2	Q9JX85	Q9jx85 neisseria m
24	4.9	91	12	O78450	O78450 human immun
25	4.9	92	12	O80035	O80035 human immun
26	4.9	94	2	O50759	O50759 borrelia bu
27	4.9	98	1	Q9HPX8	Q9hpx8 halobacteri
28	4.9	115	2	Q9RD28	Q9rd28 streptomyce
29	4.9	116	1	O58282	O58282 pyrococcus
30	4.9	116	2	O32470	O32470 aeromonas p
31	4.9	120	5	Q9U9U6	Q9u9u6 cryptospori
32	4.9	120	5	Q9TW73	Q9tw73 cryptospori
33	4.9	120	5	Q9TW72	Q9tw72 cryptospori
34	4.9	120	5	O9TW71	O9tw71 cryptospori
35	4.9	120	5	O9TVU9	O9tvu9 cryptospori
36	4.9	120	5	Q9NA06	Q9na06 drosophila
37	4.9	120	5	Q9NA05	Q9na05 drosophila
38	4.9	120	5	Q9NA04	Q9na04 drosophila
39	4.9	120	5	Q9N9Y2	Q9n9y2 drosophila
40	4.9	120	5	Q9N9Y1	Q9n9y1 drosophila
41	4.9	120	5	Q9N9Y0	Q9n9y0 drosophila
42	4.9	120	5	Q9N6H0	Q9n6h0 drosophila
43	4.9	124	7	Q08610	Q08610 aulonocara
44	4.9	142	11	Q9D7H4	Q9d7h4 mus musculu

ALIGNMENTS

RESULT: 1

O55082 ID Q55082 PRELIMINARY: PRT; 426 AA.
AC O55082;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 20 (EC 3.1.3.48)
DE (PHOSPHOTYROSINE PHOSPHATASE) (PTPASE).
GN PTPN20.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=98070510; PubMed=9407093;
RA Ohnogi M., Kuramochi S., Matsuda S., Yamamoto T.;
RT Molecular cloning and characterization of a novel cytoplasmic
RT protein-tyrosine phosphatase that is specifically expressed in
RT spermatocytes.";
RL J. Biol. Chem. 272:33092-33099(1997).
CC -; CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN
CC -; TYROSINE + PHOSPHATE.
DR EMBL; D64141; BAA23761.1; -;
DR HSSP; P28827; IRPM.
DR MCB; MG1:1196295; Ptpn20.
DR InterPro; IPR000387; TYR_phosphatase.
DR InferPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PTPPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
SQ SEQUENCE 426 AA; 49118 MW; 2B35F813379502F4 CRC64;

Query-Match 18.9%; Score 23; DB 11; Length 426;
Best Local Similarity 100.0%; Pred. No. 5.7e-17;

```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DFFIKYRVYRKSHITGPLLVC 116
Db 337 DFFIKYRVYRKSHITGPLLVC 359

RESULT 2
QY406
ID QY406 PRELIMINARY; PRT; 398 AA.
AC QY406;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 45.7 KDA PROTEIN (EC 3.1.3.48) (FRAGMENT).
GN DKFZP566K0524.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Anorge W., Winkler U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL050400; CAB43248.1; -.
DR HSSP; Q06124; 2SHP.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00102; Y_phosphatase.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 398 AA; 45690 MW; 857AAD03747870A2 CRC64;

Query Match 8.2%; Score 10; DB 4; Length 398;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KWPDHGTPAS 92
Db 298 KWPDHGTPAS 307

RESULT 3
QYTPH8
ID QYTPH8 PRELIMINARY; PRT; 52 AA.
AC QYTPH8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE MHC CLASS II ANTIGEN (FRAGMENT).
OS Xiphophorus helleri.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8084;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHCIIASWORDTAIL;
RX MEDLINE=99398697; PubMed=10468597;
RA Venkatesh B., Ning Y., Brenner S.;
RT "Late changes in spliceosomal introns define clades in vertebrate
evolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
DR EMBL; AF134946; AAD54146.1; -.
KW MHC.

```

```

FT NON_TER 1
SQ SEQUENCE 52 AA; 5839 MW; 07CFF4E11374DA1B CRC64;

Query Match 5.7%; Score 7; DB 7; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 VSLKEPL 40
Db 23 VSLKEPL 29

RESULT 4
QYD792
ID QYD792 PRELIMINARY; PRT; 116 AA.
AC QYD792;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 913001LK15RIK PROTEIN.
GN 913001LK15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Kuehl L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Kordone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009447; BAB26293.1; -.
DR MGI; MGI:1914070; 913001LK15RIK.
SQ SEQUENCE 116 AA; 13274 MW; FE2899B4A2B25EA8 CRC64;

Query Match 5.7%; Score 7; DB 11; Length 116;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 EGGVIKC 28
Db 79 EGGVIKC 85

RESULT 5
QY3103
ID QY3103 PRELIMINARY; PRT; 158 AA.
AC QY3103;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN V3-V5 LOOP REGION (FRAGMENT).

```


OS Human immunodeficiency virus type 1.
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=98160778; PubMed=9499799;
 RA Shankarappa R., Gupta P., Learn G.H. Jr., Rodrigo A.G., Ehrlich G.D.;
 RA Rinaldo C.R. Jr., Gorri M.C., Millins J.I., Nara P.L.,
 RT "Evolution of human immunodeficiency virus type 1 envelope sequences
 in infected individuals with differing disease progression profiles.";
 RL Virology 241:251-259(1998).
 DR EMBL: U15836; AAC59204.1; -.
 DR InterPro: IPR000777; GPI20.
 DR Pfam: PF00516; GPI20; 1.
 KW Envelope protein.
 FT NON_TER 1
 FT NON_TER 158
 SQ SEQUENCE 158 AA; 17431 MW; 7039619AB3EB8595 CRC64;

Query Match 5.7%; Score 7; DB 12; Length 158;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ITGPLLIV 114
 DB 142 ITGPLLIV 148

RESULT 6
 O62884 PRELIMINARY; PRT; 244 AA.

AC O62884;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MHC CLASS II BETA CHAIN (FRAGMENT).
 GN XIMA-DAB.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=JP 163 A; TISSUE=GUT;
 RA McConnell T.J., Godwin U.B., Norton S.F., Nairn R.S., Kazianis S.,
 RA Morizot D.C.;
 RL submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR
 CC HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).
 CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC -!- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,
 CC BETA-1 DOMAIN INTERPRO FAMILY.

DR EMBL: AF040760; AAC05652.1; -.
 DR HSSP: P01888; 1BMG.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR000353; MHC_II_beta.
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF00969; MHC_II_beta; 1.
 DR ProDom: PD000328; MHC_II_beta; 1.
 DR SMART: SM00407; IGc1; 1.
 KW Glycoprotein; MHC; MHC II; Transmembrane.
 FT NON_TER 1
 SQ SEQUENCE 244 AA; 27861 MW; 8B0BDC0224038580 CRC64;

Query Match 5.7%; Score 7; DB 7; Length 244;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 VSLKEPL 40
 DB 188 VSLKEPL 194

RESULT 7
 Q9XV30 PRELIMINARY; PRT; 310 AA.

AC Q9XV30;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE F40D4.5 PROTEIN.
 GN F40D4.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Matthews L.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

QY 108 ITGPLLIV 114
 DB 234 ITGPLLIV 240

RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RL EMBL: Z81536; CAB04364.1; -.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR PROSITE: PS0262; G_PROTEIN_RECEP_F1_2; 1.
 SQ SEQUENCE 310 AA; 35478 MW; FA6A37FD2A76C0C CRC64;

Query Match 5.7%; Score 7; DB 5; Length 310;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ITGPLLIV 114
 DB 234 ITGPLLIV 240

RESULT 8
 Q9SST9 PRELIMINARY; PRT; 351 AA.

AC Q9SST9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GERANYLGERANYL PYROPHOSPHATE SYNTHASE.
 OS Scoparia dulcis.
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Lamiales; Veronicaceae; Scoparia.
 OX NCBI_TaxID=107240;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kojima N., Suh D., Sankawa U.;
 RT "Scoparia dulcis GGPP synthase.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB034250; BAA86285.1; -.

DR InterPro: IPR000092; Polyprenyl_synt.
 DR Pfam: PF00348; polyprenyl_synt; 1.
 DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 SQ SEQUENCE 351 AA; 38342 MW; 247570EICD863D5C CRC64;

Query Match 5.7%; Score 7; DB 10; Length 351;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 VSLKEPL 40
 |||||
 Db 79 VSLKEPL 85

RESULT 9

Q9M7Q0 PRELIMINARY; PRT; 378 AA.
 ID Q9M7Q0;
 AC Q9M7Q0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE REPLICASE
 OS Porphyra pulchra.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
 OX NCBI_TaxID=60925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Moon D.A., Goff L.G.;
 RT "Sequence signatures of flowering plant geminiviruses are found in
 dsDNA plasmids of the evolutionarily ancient red alga Porphyra.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF106326; AAF36422.1; -
 DR InterPro: IPR001191; Gemini.AL1.
 DR Pfam: PD000736; Gemini.AL1; 1.
 DR ProDom: PD000736; Gemini.AL1; 1.
 SQ SEQUENCE 378 AA; 43495 MW; 056A183EA635FDC1 CRC64;

Query Match 5.7%; Score 7; DB 10; Length 378;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HFHVLE 50
 |||||
 Db 62 HFHVLE 68

RESULT 10

Q9M7P9 PRELIMINARY; PRT; 378 AA.
 ID Q9M7P9;
 AC Q9M7P9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE REPLICASE.
 OS Porphyra pulchra.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
 OX NCBI_TaxID=60925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Moon D.A., Goff L.G.;
 RT "Sequence signatures of flowering plant geminiviruses are found in
 dsDNA plasmids of the evolutionarily ancient red alga Porphyra.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF106327; AAF36423.1; -
 DR InterPro: IPR001191; Gemini.AL1.
 DR Pfam: PF00799; Gemini.AL1; 1.
 DR ProDom: PD000736; Gemini.AL1; 1.
 SQ SEQUENCE 378 AA; 43680 MW; 1C8E9E4A9E68EE14 CRC64;

Query Match 5.7%; Score 7; DB 10; Length 378;

Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 HFHVLE 50
 |||||
 Db 62 HFHVLE 68

RESULT 11

Q9M7P8 PRELIMINARY; PRT; 378 AA.
 ID Q9M7P8;
 AC Q9M7P8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE REPLICASE.
 OS Porphyra pulchra.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
 OX NCBI_TaxID=60925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Moon D.A., Goff L.G.;
 RT "Sequence signatures of flowering plant geminiviruses are found in
 dsDNA plasmids of the evolutionarily ancient red alga Porphyra.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF106328; AAF36424.1; -
 DR InterPro: IPR001191; Gemini.AL1.
 DR Pfam: PF00799; Gemini.AL1; 1.
 DR ProDom: PD000736; Gemini.AL1; 1.
 SQ SEQUENCE 378 AA; 44047 MW; FFE240CDBF4AEFE0 CRC64;

Query Match 5.7%; Score 7; DB 10; Length 378;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HFHVLE 50
 |||||
 Db 62 HFHVLE 68

RESULT 12

Q9XV85 PRELIMINARY; PRT; 397 AA.
 ID Q9XV85;
 AC Q9XV85;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE F20E11.2 PROTEIN.
 DE F20E11.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ainscough R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";

```
RL Nature 368:32-38(1994).
DR EMBL: Z81508; CAB04140.1; -.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR PROSITE; PS50262; G-PROTEIN_RECEP_FL_2; 1.
SQ SEQUENCE 397 AA; 44724 MW; 7BE8E2FFEE36EC4 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26; Length 397;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 ITGPLLV 114
      |||||
Db 234 ITGPLLV 240

RESULT 13
Q9AA59 PRELIMINARY; PRT; 407 AA.
AC
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OMPA FAMILY PROTEIN.
GN CC0747.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OC NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005750; AAK22732.1; -.
DR TIGR: CC0747; -.
KW Complete proteome.
SQ SEQUENCE 407 AA; 42781 MW; 54610D5FA64CC3BF CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 27; Length 407;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 CTAGVGR 122
      |||||
Db 105 CTAGVGR 111

RESULT 14
Q9Y1X4 PRELIMINARY; PRT; 446 AA.
AC
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SPTR5 (EC 3.1.3.48) (FRAGMENT).
OS Ephydatia fluviatilis.
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Haplosclerida; Spongillidae; Ephydatia.
OC NCBI_TaxID=31330;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99246376; PubMed=10229569;
RA Ono K., Suga H., Iwabe N., Kuma K., Miyata T.;
RT "Multiple protein tyrosine phosphatases in sponges and explosive gene
```

```
RT duplication in the early evolution of animals before the parazoan-
RT eumetazoan split."
J. Mol. Evol. 48:654-662(1999).
DR EMBL: AB019127; BAA82560.1; -.
DR HSSP; P18052; 1VFO.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON_TER
SQ SEQUENCE 446 AA; 50365 MW; 9E5B8AF6168FDF7C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 29; Length 446;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 GPLLVHC 116
      |||||
Db 152 GPLLVHC 108

RESULT 15
Q9NL12 PRELIMINARY; PRT; 488 AA.
AC
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE AMPTPR5 PROTEIN (FRAGMENT).
GN AMPTPR5.
OS Branchiostoma belcheri.
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OC NCBI_TaxID=7741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219325; PubMed=10754074;
RA Ono-Koyanagi K., Suga H., Katoh K., Miyata T.;
RT "Protein tyrosine phosphatases from amphioxus, hagfish, and ray:
RT divergence of tissue-specific isoform genes in the early evolution of
RT vertebrates."
J. Mol. Evol. 50:302-311(2000).
DR EMBL: AB033565; BAA95172.1; -.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON_TER
SQ SEQUENCE 488 AA; 55795 MW; A3DB8EB912DD7226 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 31; Length 488;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 WPDHGTP 90
      |||||
Db 78 WPDHGTP 84

Search completed: March 30, 2002, 08:31:09
```

Job time: 261 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:20:42 ; Search time 1468.71 Seconds
(without alignments)
3594.377 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAAATATTTGTAATGTAT.....ATTACAGACCCCTCTTGT 320

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo.hum.*
- 31: em.htgo.inv.*
- 32: em.htgo.rod.*
- 33: em.htg.hum.*
- 34: em.htg.inv.*
- 35: em.htg.rod.*
- 36: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	280	87.5	3090	10	D64141	D64141 Mus musculus
2	241.6	75.5	2786	9	HSM800374	AL050040 Homo sapi
3	144.8	45.3	197389	9	AL358791	AL358791 Human DNA
4	57	17.8	4424	9	HSPNP1	X79676 H.sapiens P
5	57	17.8	7546	9	HUMTPB3	D21211 Human mRNA
6	57	17.8	8040	6	AR048218	AR048218 Sequence
7	57	17.8	8043	9	HSPTP1	X80289 H.sapiens P
8	57	17.8	8062	9	HUMTPB2	D21210 Human mRNA
9	57	17.8	8119	6	AR143153	AR143153 Sequence
10	57	17.8	8119	9	HUMTPB1	D21209 Human mRNA
11	57	17.8	8287	9	HSU12128	U12128 Human prote
12	52.8	16.5	8411	4	BTU20807	U20807 Bos taurus
13	49.6	15.5	2433	9	BC010863	BC010863 Homo sapi
14	49.6	15.5	3956	9	HUMTPSA	M83738 Human prote
15	47.2	14.8	4756	9	HUMTPS	D37781 Human mRNA
16	47.2	14.8	5117	6	AR109903	AR109903 Sequence
17	47.2	14.8	5117	9	HSU10886	U10886 Human densi
18	46.6	14.6	2676	10	MMPTPPES	X63440 M.musculus
19	46.6	14.6	2983	10	NMP19PTP	U00638 Human mRNA
20	46	14.4	6314	10	RNU40790	U40790 Rattus norv
21	45.6	14.2	4315	9	HSLCAR	Y00622 Human mRNA
22	45.6	14.2	4597	9	HSICA	AF150732 Homo sapi
23	45	14.1	2877	9	AF150732	AX020263 Sequence
24	44.2	13.8	2356	6	AX020263	AF001847 Homo sapi
25	44.2	13.8	2356	9	AF001847	AX020261 Sequence
26	44.2	13.8	3058	6	AX020261	AF001846 Homo sapi
27	44.2	13.8	3058	9	AF001846	AF198450 Xenopus lae
28	44.2	13.8	3472	5	AF198450	AR149916 Sequence
29	44.2	13.8	3580	6	AR149916	D45414 Rat mRNA fo
30	43.4	13.6	2321	10	RATBEM3	AF332084 Mus muscu
31	43.4	13.6	2674	10	AF332084	AF332083 Mus muscu
32	43.4	13.6	2683	10	AF332083	X92563 R.norvegicu
33	43.4	13.6	2952	10	RNICA105	A37107 Sequence 1
34	43.4	13.6	3396	6	A37107	U40652 Rattus norv
35	43.4	13.6	3397	10	RNU40652	U11812 Mus musculu
36	43.4	13.6	3446	10	MMU11812	D38222 Rat mRNA fo
37	43.4	13.6	3477	10	RATPDPTPLP	A37109 Sequence 3
38	43.4	13.6	3561	6	A37109	X74438 M.musculus
39	43.4	13.6	3566	10	MMPTP35A	AF013490 Mus muscu
40	43.2	13.5	3531	10	AF013490	LI5420 Dictyostell
41	43	13.4	1731	3	DDIPTPA	L33099 Xenopus lae
42	43	13.4	1863	5	XELTYPHA	AC073272 Homo sapi
43	43	13.4	54594	9	AC073272	Z23060 M.musculus
44	42.8	13.4	318	10	NMPRTYPHM	AB033587 Potamo
45	42.4	13.2	1812	5	AB033587	

ALIGNMENTS

RESULT 1

D64141

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

1

D64141

Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.

3090 bp

mRNA

13-FEB-1999

ROD

GI:2665457

protein-tyrosine-phosphatase.

Mus musculus testis cDNA to mRNA.

Mus musculus

1 (bases 1 to 3090)

Ohsugi, M.

Direct Submission

Submitted (15-SEP-1995) to the DDBJ/EMBL/GenBank databases. Miho

Ohsugi, The University of Tokyo, Department of Oncology; 4-6-1

Shirokanedai, Minato-ku, Tokyo 108, Japan

(E-mail: mhsugi@ims.u-tokyo.ac.jp, Tel: 03-5449-5305,

Fax: 03-5449-5413)

2 (sites)

AUTHORS Ohsui, M., Kuramochi, S., Matsuda, S. and Yamanoto, T.
TITLE Molecular cloning and characterization of a novel cytoplasmic protein-tyrosine phosphatase that is specifically expressed in spermatocytes
JOURNAL J. Biol. Chem. 272 (52), 33092-33099 (1997)
MEDLINE 98070510
FEATURES
source
1. .3090
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="testis"
218. 1498
/EC_number="3.1.3.48"
/codon_start=1
/product="protein-tyrosine-phosphatase"
/protein_id="BAA23761.1"
/db_xref="GI:2665458"
/translation="MSSPRKVRKGTGRDNDDEEGNSCNLNLNSLPSSQKMTPTKPI FGKMSNVKPSHLHSFSDKYELVYPEPESDTDETVDSDRSURNWNSMDSETA GPKSTVPLVSGSSRLSKDSTYSKELTQLAQIRPLFNFSARSAMRDCLNLTQKK EELDIITRELEQMTLPDQNSGNTLQNRKNRDIPLDYSTRVPLGKNKDYINAS YIRIVNHEBEYFIATQGPLPETIEDFWQVLENNCNVIAITRETCGVKCYSWP ISLKEPLEFHSVLETFHTVTOYFTVRVFOIVKSTGKSCQVKHLQFTKWPDPGTPA SADFFIKYRVYVKSHTLGPDLVHCSAGVGRGVFCVDVWFSAEKNSYFDLMNIVT QMRKQCGMLQTKEQYQFCVEIVLEVLQNLALY"
BASE COUNT 904 a 603 c 647 g 936 t
ORIGIN
Query Match 87.5%; Score 280; DB 10; Length 3090;
Best Local Similarity 92.2%; Pred. No. 8.2e-68;
Matches 299; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 1 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAGTGGAGTTATCAAG 60
Db 968 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAGTGGAGTTATCAAG 1027
QY 61 TGTTCAGATTACTGGCCGCTTCTCTGAAGAGCCTTTGGAATTCACAACTTCATGTC 120
Db 1028 TGTTCAGATTACTGGCCGCTTCTCTGAAGAGCCTTTGGAATTCACAACTTCATGTC 1087
QY 121 CTCTCGAGAACTTCAGATAACACAGTATTTGTCATCGAATATTCAAATTTGTAAG 180
Db 1088 TTCTCGAGAACTTCAGATAACACAGTATTTGTCATCGAATATTCAAATTTGTAAG 1147
QY 181 AAGTCCACAGAAAGAGTACTCTGTAAACACTGTGAGTTCAATCAATGCCAGACCAT 240
Db 1148 AAGTCCACAGAAAGAGTACTCTGTAAACACTGTGAGTTCAATCAATGCCAGACCAT 1207
QY 241 GGCACTCTGCCTCAGTAGATTTTTCATCAATATATGTCGTTATGTCGAGAGAGCCAC 300
Db 1208 GGCACTCTGCCTCAGTAGATTTTTCATCAATATATGTCGTTATGTCGAGAGAGCCAC 1267
QY 301 ATTACAGACCCCTCCTTGT 320
Db 1268 ATTACAGACCCCTCCTTGT 1287
RESULT 2
HSM800374 2786 bp mRNA PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFZp566K0524 (from clone DKFZp566K0524);
DEFINITION partial cds.
ACCESSION AL050040
VERSION AL050040.1 GI:4884281
KEYWORDS human,
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2786)
AUTHORS Ansoorge, W., Wirkner, U., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission

JOURNAL Submitted (15-MAY-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp566K0524) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1. .2786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp566K0524"
/clone_1b="566 (synonym: hfk2). Vector pAMP1; host XL-2blue; sites NotI + SalI"
/dev_stage="fetal"
/tissue_type="kidney"
1. .1198
/gene="DKFZp566K0524"
/c1..1198
/gene="DKFZp566K0524"
/note="strong similarity to protein-tyrosine-phosphatases"
/codon_start=2
/product="hypothetical protein"
/protein_id="CAB43248.1"
/db_xref="GI:4884282"
/translation="GNWRGCGWCGTGAAPQAVFENKVNSEKVLRLNPPHNDY EDVFEPSGSDPSMTARGPFRDRWSEDEAAGPQAALSPILSDTRKIVSEGL DOLAQRPLIFNFHEQTAKDCLILEETAAADYIMQERMALEKLNLPGEFYSGNPS NREKNRYDILPYDSTRVPLGSKDYNASYIRVNGEYFYATQGPIISIDDFW QMVLNNSNVIAITREMEGGIICKYVWPLSLKPLELKHFFLENVQLLOYFYLR MQQVKEKSTGTSYKQLQFTKPDHGTASADSFYKIRYARKSHLTGPMVHCSAG IGTGVFLCLVDVVFCAIVKDCSFNIMDIVAQMREKRSQWVQTEQTHFCYDVLVLR KLLTLD"
polyA_signal 2761. .2766
polyA_site 2776
BASE COUNT 841 a 494 c 550 g 901 t
ORIGIN
Query Match 75.5%; Score 241.6; DB 9; Length 2786;
Best Local Similarity 84.7%; Pred. No. 4.3e-57;
Matches 271; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 1 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAGTGGAGTTATCAAG 60
Db 668 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAGTGGAGTTATCAAG 727
QY 61 TGTTCAGATTACTGGCCGCTTCTCTGAAGAGCCTTTGGAATTCACAACTTCATGTC 120
Db 728 TGTTCAGATTACTGGCCGCTTCTCTGAAGAGCCTTTGGAATTCACAACTTCATGTC 787
QY 121 CTTCGAGAACTTCAGATAACCTCAGTATTTGTCATCGAATATTCAAATTTGTAAG 180
Db 788 TTCTCGAGAACTTCAGATAACCTCAGTATTTGTCATCGAATATTCAAATTTGTAAG 847
QY 181 AAGTCCACAGAAAGAGTACTCTGTAAACACTGTGAGTTCAATCAATGCCAGACCAT 240
Db 848 AAGTCCACAGAAAGAGTACTCTGTAAACACTGTGAGTTCAATCAATGCCAGACCAT 907
QY 241 GGCACTCTGCCTCAGTAGATTTTTCATCAATATATGTCGTTATGTCGAGAGAGCCAC 300
Db 908 GGCACTCTGCCTCAGTAGATTTTTCATCAATATATGTCGTTATGTCGAGAGAGCCAC 967
QY 301 ATTACAGACCCCTCCTTGT 320
Db 968 ATTACAGACCCCTCCTTGT 987

RESULT 3

AL358791/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-42B19 on chromosome 10, complete sequence.
 ACCESSION AL358791
 VERSION AL358791.24 GI:14141350
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 197389)
 Clark,S.
 Direct Submission
 Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On May 17, 2001 this sequence version replaced gi:13751420.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
 RP11-42B19 is from the library RCL11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-42B19.

FEATURES
 source
 1..197389
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-42B19"
 /clone.lib="RCL1-11.1"
 6..370
 /note="match: GSS: Em:AQ116178"
 26..180
 /note="match: GSS: Em:AQ116242"
 76..566
 /note="match: GSS: Em:AQ726140"
 1049..1114
 /note="6 copies 11 mer 83% conserved"
 1192..1246
 /note="5 copies 11 mer 87% conserved"
 1246..1660
 /note="L2 repeat: matches 2256..2667 of consensus"
 2964..3011
 /note="L1Me repeat: matches 575..621 of consensus"
 3170..3580
 /note="MSTA repeat: matches 1..426 of consensus"
 4275..4345
 /note="MIR repeat: matches 106..181 of consensus"
 5259..5374
 repeat_region

repeat_region
 5414..5454
 /note="MIR repeat: matches 111..227 of consensus"
 repeat_region
 5633..6052
 /note="L2 repeat: matches 2657..2697 of consensus"
 repeat_region
 6053..6357
 /note="LTR12 repeat: matches 1..342 of consensus"
 repeat_region
 6358..6450
 /note="AluSq repeat: matches 1..304 of consensus"
 repeat_region
 9115..9246
 /note="LTR12 repeat: matches 342..440 of consensus"
 repeat_region
 9122..9261
 /note="12 copies 11 mer 65% conserved"
 repeat_region
 11013..11192
 /note="35 copies 4 mer aaga 66% conserved"
 repeat_region
 11231..11638
 /note="MER45 repeat: matches 1..178 of consensus"
 misc_feature
 11616..12203
 /note="match: GSS: Em:AQ210526"
 repeat_region
 12204..12435
 /note="MER48 repeat: matches 1..611 of consensus"
 repeat_region
 12452..12704
 /note="L2 repeat: matches 2251..2500 of consensus"
 repeat_region
 12796..12883
 /note="MIR repeat: matches 2..262 of consensus"
 repeat_region
 12939..13288
 /note="L2 repeat: matches 2641..2734 of consensus"
 misc_feature
 13577..13719
 /note="match: GSS: Em:A2694508"
 repeat_region
 13747..13866
 /note="FLAM_C repeat: matches 1..143 of consensus"
 repeat_region
 13895..14079
 /note="MER93 repeat: matches 110..230 of consensus"
 repeat_region
 14090..14220
 /note="MER93 repeat: matches 215..397 of consensus"
 repeat_region
 14657..14971
 /note="MER5B repeat: matches 25..178 of consensus"
 repeat_region
 15219..16945
 /note="AluSc repeat: matches 1..309 of consensus"
 repeat_region
 16946..17256
 /note="L1MD repeat: matches 1905..3500 of consensus"
 repeat_region
 17257..18226
 /note="AluY repeat: matches 1..311 of consensus"
 repeat_region
 18902..19076
 /note="L1MD repeat: matches 977..1905 of consensus"
 repeat_region
 20392..20519
 /note="L1MD repeat: matches -15..174 of consensus"
 repeat_region
 20677..20808
 /note="MLT11 repeat: matches 6..129 of consensus"
 repeat_region
 20833..20868
 /note="FLAM_C repeat: matches 2..133 of consensus"
 repeat_region
 20869..21214
 /note="MLT11 repeat: matches 308..346 of consensus"
 repeat_region
 21217..21621
 /note="MLT1A1 repeat: matches 1..365 of consensus"
 repeat_region
 22088..22592
 /note="MLT1A1-internal repeat: matches 927..1333 of consensus"
 repeat_region
 22617..22732
 /note="MLT1A1-internal repeat: matches 160..712 of consensus"
 repeat_region
 22795..22911
 /note="L1ME3 repeat: matches 6022..6145 of consensus"
 repeat_region
 22917..23273
 /note="MLT1A1-internal repeat: matches 5..121 of consensus"
 repeat_region
 23274..23344
 /note="MLT1A1 repeat: matches 1..365 of consensus"
 repeat_region
 24336..24385
 /note="MLT11 repeat: matches 339..410 of consensus"
 repeat_region
 25059..26364
 /note="MLTIF repeat: matches 268..317 of consensus"
 repeat_region
 26398..26843
 /note="L1M4 repeat: matches 1064..2552 of consensus"

Matches	163;	Conservative	0;	Mismatches	155;	Indels	3;	Gaps	1;
Qy	1	GAATAAATTTGTAATGTTATTGCTATGATAACACAGAGAGATAGAAGGTGGAGTTATCAAG	60						
Db	6901	GAGCAAAATCCACAGTGAATGACCATGATGACTCAAGAAGTGAAGGAGAGAAAAATCAA	6960						
Qy	61	TGTTGCAGTTACTGGCCGGTTTCTCT---GAAGGAGCCCTTTTGGAAATTCAAACACTTTTCAT	117						
Db	6961	TGCCAGCGCTATTGGCCCAACATCCTAGGCAAAACAACAATGGTCAGCAACAGACTTCGA	7020						
Qy	118	GTCCCTTCGGAGAACTTTTCAGATAACTCAGTATTTTGTGCATCCGAATATTTCAAAATTCGTG	177						
Db	7021	CTGGCTCTTTGTGAGAAATGCACAGCTGAAGGGCTTTTGTGTGAGGGCAATGACCCCTTGA	7080						
Qy	178	AAGAAGTCCACAGGAAGAGTCACTCTGTAAACACTTTCGAGTTTCATCAAAATGGCCAGAC	237						
Db	7081	GATATTACAGACAGAGAGGTGGCCATATTTCTCATCTGAATTTTCACATGCGCTGGCCAGAC	7140						
Qy	238	CATGGCACTCGCTCAGTAGATTTTTTTCATCAATATGTCGGTTATGTGAGGAAGAC	297						
Db	7141	CATGATACACCTTCTCAACAGATGATCTGCTTACTTTTATCTCCTACATGAGACATC	7200						
Qy	298	CACATTACAGGACCCCTCCTT	318						
Db	7201	CACAGATCAGGCCCAATCATTT	7221						
RESULT	9								
AR143153									
LOCUS	AR143153	8119 bp	DNA	PAT	08-AUG-2001				
DEFINITION	Sequence 45 from patent US 6204055.								
ACCESSION	AR143153								
VERSION	AR143153.1	GI:15104439							
KEYWORDS	Unknown.								
SOURCE	Unknown.								
ORGANISM	Unclassified.								
REFERENCE	1 (bases 1 to 8119)								
AUTHORS	Dean, N.M. and Marcussen, E.G.								
TITLE	Antisense inhibition of Fas mediated signaling								
JOURNAL	Patent: US 6204055-A 45 20-MAR-2001;								
FEATURES	Location/Qualifiers								
source	1..8119								
BASE COUNT	2597 a 1674 c 1742 g	2106 t							
ORIGIN	/organism="unknown"								
Query Match	17.88;	Score 57;	DB 6;	Length 8119;					
Best Local Similarity	50.88;	Pred. No. 1.6e-05;							
Matches	163;	Conservative	0;	Mismatches	155;	Indels	3;	Gaps	1;
Qy	1	GAATAAATTTGTAATGTTATTGCTATGATAACACAGAGATAGAAGGTGGAGTTATCAAG	60						
Db	6958	GAGCAAAATCCACAGTGAATGACCATGATGACTCAAGAAGTGAAGGAGAGAAAAATCAA	7017						
Qy	61	TGTTGCAGTTACTGGCCGGTTTCTCT---GAAGGAGCCCTTTTGGAAATTCAAACACTTTTCAT	117						
Db	7018	TGCCAGCGCTATTGGCCCAACATCCTAGGCAAAACAACAATGGTCAGCAACAGACTTCGA	7077						
Qy	118	GTCCCTTCGGAGAACTTTTCAGATAACTCAGTATTTTGTGCATCCGAATATTTCAAAATTCGTG	177						
Db	7078	CTGGCTCTTTGTGAGAAATGCACAGCTGAAGGGCTTTTGTGTGAGGGCAATGACCCCTTGA	7137						
Qy	178	AAGAAGTCCACAGGAAGAGTCACTCTGTAAACACTTTCGAGTTTCATCAAAATGGCCAGAC	237						
Db	7138	GATATTACAGACAGAGAGGTGGCCATATTTCTCATCTGAATTTTCACATGCGCTGGCCAGAC	7197						
Qy	238	CATGGCACTCTCGCTCAGTAGATTTTTTTCATCAATATGTCGGTTATGTGAGGAAGAC	297						
Db	7198	CATGATACACCTTCTCAACAGAGATGATCTGCTTACTTTTATCTCCTACATGAGACATC	7257						
Qy	298	CACATTACAGGACCCCTCCTT	318						

```

Db      7258      CACAGATCAGGCCCAATCATT 7278
|||||  |||||  |||||  |||||
RESULT  10
HUMPTPB1
LOCUS   Human mRNA for protein tyrosine phosphatase (PTP-BAS, type 1),
DEFINITION complete cds.
ACCESSION D21209.1 GI:452189
KEYWORDS PTP-BAS; protein tyrosine phosphatase.
SOURCE    Homo sapiens (library: Lambda Zap II) basophilic leukemia cells
          cell-line K0812E cDNA to mRNA, clones BAS-[1-4].
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS  Maekawa,K.
TITLE     Direct Submission
JOURNAL   Submitted (15-OCT-1993) to the DDBJ/EMBL/GenBank databases.
          Kazuhiko Maekawa, Shionogi Institute for Medical Science; 2-5-1
          Mishima, Settsu, Osaka 566, Japan (Tel:06-382-2612,
          Fax:06-382-2598)
          2 (bases 1 to 8119)
          Molecular cloning of a novel protein-tyrosine phosphatase
          containing a membrane-binding domain and GLGF repeats
          FEBS Lett. 337 (Oct. 1994) 94116679
          Submitted (15-Oct-1993) to DDBJ by:
          Kazuhiko Maekawa
          Shionogi Institute for Medical Science
          2-5-1 Mishima, Settsu
          Osaka 566
          Japan
          Phone: 06-382-2612
          Fax: 06-382-2598
FEATURES
source
1..8119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="K0812E"
/cell_type="basophilic leukemia cells"
/clone_lib="Lambda zap II"
64..7521
/codon_start=1
/product="protein tyrosine phosphatase (PTP-BAS, type 1)"
/protein_id="BAA04750.1"
/db_xref="GI:452190"
/translation="MHVSLAEALVEVRGGLPQEEIEIWAVLNQSAESLQELFRKVSADP
AALGFTIIPSWLLDLPSSGVSTFDENISNODLRAFTAPEVLQNOISLTSLSVDVEKIHV
NCGMTLYWGADVEVPQSQPTKGLDHLSILGMCEDVIYARVSVRTLACSAHINS
SLCAPTSYVKVHLVLLKLNLSCTDOLSCNEQKPDPSQAIRLGRKLGKGLPTGRSSDTS
VLDIQAPLSHQFTLNNGLSKSMGFLSIKQDQENFKDILNSGREDENTFSPIQ
FKTSGPKKPIPGIDVLNKKKIWAQSMDDLCTADRFSSGATATYRRCHPEAVTVTS
FTTPRKEARYSDGSIADLIFGQKMDPIYHRELPTSSATISALDRIREROKKLQVL
EANNVEPVRKYTYHGDVFTSSSEPSIISSSDFRQVRRSSEAKRFSSSGSLPGVD
ETLSQCSQSPRSQYETPEGNLINOEIMLKROEELMOLQAQMARLQRLSLYPGDT
IKASMLDITRDLPURELTAETQRLNFFGPEPVKMTTEPTISLDLPRSLTKKGG
NEDNRKKNMILLNGQRLTCTDKTICKDVPDMVVAHGLVEHFLPALATLKONEYF
FVDPDLKLTVAPEGKKEPKKTKATVNTLFFRIKPFMDVSLQHTLTCHQYLYLQ
LRKDIILEERHCHDTSLLASLALQAEYGDQYGVHVSYSFMEHYLFPARVMEKLDL
SYTKLEPKLHNTYVGASEKETLELFKVCQDTEYGVHFRVHPEPKKQSTGILLGV
SKGLVYFVHNGVRTLVLFRPMRETKKISFSSKKITLQNTSDGIKHGFTDNSKIQCY
LHLCYSQGFOLQMARQSNODADIERASPSNLQAEYSVRGFMNGRAISTGSLAS
STLNKLAVRPLSWVMEKPLKSLSELSQLQPNSSKEKNDKASWEEKPREMKSVASL
LSCASLYPIRKWYVNMPEPPQTVAELVGKPSHQMSRSDAESLAGYTKLNNKSVASL
NRSPEARLKHEDSSSDTEGQQATVLGTMHSSGNSQVPLKENDVILHRKWSIVSPSE
REITLVNKKDAGKYLGFOLIGENKMRGLDGIFTSSVAPGGPADGGCLKPGEDRLIS
VNSVLEGVSHHAAITLQNAEDPTLVLSQPEKTSKYVSTPVIHLENTNMKNYMKSS
YMQDSALSSDKDHWMSRGTLRHISSENSFGSGLREGSLSDQSRTESASLSQSOVN
GPFASHLGDTWOESQGSPPSVISKATEKETFTDTSNOSKTKKPGISDVTDYSDRGV

```

[illegible]

Query Match	17.8%	Score 57	DB 9	Length 8119	
Best Local Similarity	50.8%	Pred. No. 1.6e-05			
Matches 163	Conservative	0	Mismatches 155	Indels 3	Gaps 1

QY	1	GAAATAATTCTAATGTATTGTCTATGATACACAGAGAGATAGAGGTGGAGTTATCAAG	60
Db	6958	GAGCAAAATCCACAGATAGCCCATGATGACTCAAGAAGTAGAAGGAGAAAAATCAAA	7017
QY	61	TGTTGCAGTTACTGGCCCGTTTCTCTCT---GAAGGAGGCTTTGGAAATTCAAACACTTTTCAT	117
Db	7018	TGCCAGCGCTATTGGCCCCAACATCCTAGGCCAAACAACAATGGTCAGCAACAGACTTCGA	7077
QY	118	GTCCCTTCGTGGAGAACTTTCAGATAACTCAGATATTTTGTGTCATCCGAATATTTCAAAATGTG	177
Db	7078	CTGCTCTTGTGAGATGCGACAGCTGAAGGGCTTTGTGGTGAGGGCAATGACCCTTGAA	7137
QY	178	AAGAAGTCCACAGAGAAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAAAATGSCCAGAC	237
Db	7138	GATATTCAGACCGAGAGGTGGCCATATTTCTCATCTGAAATTTCACTGCGTCGCCAGAC	7197
QY	238	CATGGCACTCTCTGCTCAGTAGATTTTTTTCATCAAAATATGTCCGTTATGTGAGAAAGAGC	297
Db	7198	CATGATACACCTTCTCAACAGATGATCTGCTTACTTTTATCTCTCATGAGACATC	7257
QY	298	CACATTCAGAGACCCCTCCCTT	318
Db	7258	CACAGATCAGGCCCAATCAAT	7278

RESULT	11
LOCUS	HSU12128
LOCUS	HSU12128 8287 bp mRNA PRI 12-OCT-1994
DEFINITION	Human protein tyrosine phosphatase 1E (PTP1E) mRNA, complete cds.
ACCESSION	U12128
VERSION	U12128.1 GI:557287
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 8287)
AUTHORS	Banville,D., Ahmad,S., Stocco,R. and Shen,S.H.
TITLE	A novel protein-tyrosine phosphatase with homology to both the

JOURNAL	cytoskeletal proteins of the band 4.1 family and
MEDLINE	junction-associated guanylate kinases
REFERENCE	J. Biol. Chem. 269 (35), 22320-22327 (1994)
AUTHORS	94350988
TITLE	2 (bases 1 to 8287)
JOURNAL	Banville, D.
COMMENT	Direct Submission
FEATURES	Submitted (08-JUL-1994) Denis Banville, Biotechnology Research Institute, Pharmaceutical Biotechnology, 6100 Royalmount, Montreal, Quebec H4P 2R2, Canada
JOURNAL	3 (bases 1 to 8287)
MEDLINE	Maekawa, K., Imagawa, N., Nagamatsu, M. and Harada, S.
REFERENCE	Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane-binding domain and GLGF repeats
AUTHORS	FEBS Lett. 337 (2), 200-206 (1994)
TITLE	94116679
JOURNAL	On Oct 12, 1994 this sequence version replaced g1:515682.
MEDLINE	Location/Qualifiers
COMMENT	1..8287
FEATURES	source

```

5'UTR
<1..217
/clone_11b="ZR-75-1 breast carcinoma cDNA library"
/notes="nucleotide sequence 1-212 is one of five types of
hPTP1E 5' non-coding sequences identified by 5' RACE PCR"
/evidence=experimental
218..7690
/gene="PTP1E"
CDS
218..7690
/gene="PTP1E"
;
/notes="similar to PTP-BAS1 gene, GenBank Accession Number
D21209"
/codon_start=1

```

/protocol_id= AAB060339.1
 /db_ref="GI: 557288"
 /translation="MHVSLAEALLEVRRGGPLQEEEIWAIVLNSQASAEQLSELRVSKLADP
 AALGFIISPSWLLLPSSGVSTDENISNODLRAFTAPEVFNQSLSTSDVKBHIIY
 SLGMTLVWADYEVPOSQPIKIGDHLNLSLIGMCEDVIARVSRTVLQDACSIAHNS
 NCAPFSYVHLVKLVGLNLSGDTDLQSLCSHSEQPDRSQAIRDLRKGKGLPTGRSSSTSD
 VLIDQKPPUSHOTFLNKGUSKSGWFLUSIKDQDENFTKDLSDNSGREDSESTPSYQ
 FTKSPKEARYSDGISLADIKWASMDLITADROFSSETATYRRCHEPAVTVRTS
 TTPRKEARYSDGISLADIFGQKMDPIVHTRELPTSSAISALDRIRBROKQLOVLR
 EAMNVEEPVRKYTHGDFVSTSSSEPSLISSESDPQVRRRSPASKRFPSSCSGLPQD
 ETLQSOGSORPSQRYETPEGNILNOEIMLKQEEELMQLQAKMLRQKRLSLTPGDT
 IKASMDIPLTQURELAEATAMTORKRLNFGFGEVFKMTIEPFLISDLPSRLTKKGK
 NEDNRKVNIMLLNGORLELCTDKTICKDFDMQVHAHLGVEHLLFALATLKDNYFL
 FVDPDLQVLAPEGWKEGPKKTKATVNTFLFRIKFFMDVSLTQHTLTCHQYVLO
 LKRLDLEERHCHDDETSLLASLALQAEYQDQPEVGHVSYPFMEHYLPARVMEKLDL
 SYIKELPLKHTNYVGAASKETLEFLKVCQRLTEYGVHFRVHVIPEKKSQTGLLGV
 SKGLVFEVHNGVTVLIPFWRHETKLSFSKKKTLQNTSDGIKHGFTQDNSKICQY
 LHLHCSQHKFQOLMAROSNDOADIERASFLRNLQAEVRGFNMGRAISTGSLAS
 SYTLNKLAFLPSQABILKRLKSLSELSPQLONSKEKNDKASKEPREMKSSVYHD
 LSOASLYPHRKVNVNMEPPQTVAEVLCKSHOMSRSDAESLAGVTKLNNSKVASL
 NRSPPRKHKSUSSTIEDGQOAYVLGKTHHSGSNSSQVPLEKNDVHLKRWLSVSPPE
 RETILVNLKSDAKYGLGFQIIGEGKMRDLGLIFTSVAPGPAADDLGLKPGDLRLIS
 VNSVLEGVSHHAAIILIQNAEDPTVLVSQPKELISKVSTPVHLTNEKMTNKKFS
 YAMSDAISSSDKHWSRGTLRHISENSFGPGLRGSLSDSSTESLASOSOVN
 GFASFHLQDQWQESOHGSPSPSVISKATEKETETDSNOSKTKPGIGSDVTDYSDRGD
 SMDRETYSQQDHQTPQOESSSSVNTSKMNFKTFSSPPKPGDITVELAKNNSL
 GLSVTVLFQDQSVSVRHHGVIKAVIPQGAESDGRHIIKHGDRVLAVNGVSLGEGH
 KOAVETLRNTGVQVHULLBKPGOSTSKEHVPTPQCLSDQNAQGGQPEKRVKKTQVK
 DYSTPEENTFEVKLPKNSGSLGFSREDNLIPEQINASIVRVKLFPQAPAAESGK
 IDQGVILAVNGASLKSQOEVIISALRGTAPEVFLILCRPPGLPEIDTALLTPLO
 SPAQVLPNSKOSQSPQVQSTSDENEMSDKSKKQCSPPRRDSYSSGSGEDDL
 VAPLNPINSWSLSHQTLNSMWQAASHIEAPKQSDICTIMFYYPQIPKNPFE
 DNPSPPLDMPAQSPQOSSEASSSMDKHIIHISEPTROENMTPLKNDLENLHE
 DEFEVELLITILIKSEKSGSLFTVTKGNRIGCYVHVDIQQDPAKSGRLKPGDLRIK
 NDTQVTNMTHTDAVNLLRAASKTVLIGRVELPRIPMLPHLIPDITLTCNKEELGF
 SLCSGHDSLQWVYISDINPRASVALEGNLLQDLVDIHYNGVSTQGMTLEEVNRLDM
 LPPVSSKRSASVAGNLSLTKNGSVSYGSCSQAPLSTDSFSTVAG
 EETINETSYPKGCSTGYIKGSPNLTPKESYIODEDDIYQDSEAEVYSQSLVDVWAG
 ONLLNEETASCSGPTGLKMGKLSERTEDTDGSPLEPYETFEATKMGCEYVFC
 ONLLNEETASCSGPTGLKMGKLSERTEDTDGSPLEPYETFEATKMGCEYVFC

[illegible]

Db	7463	GATATTACAGCGGAGGAGTACGACATGTTTCTCATCTGAATTTCACTGCTGTGCCAGAC	7522
Qy	238	CATGGCACTCCTGCTCAGTAGATTTTTTTCATCAAAATATGTCGCTTATGTGAGGAAGAGC	297
Db	7523	CACGATACACTTCCAGCAGCAGCAGCAGCTGTGTGACATTTATCTCCTATATGAGACATGTT	7582
Qy	298	CACATTACAGAGACCCCTCCT	317
Db	7583	CACAGATCGGGCCCAATCAT	7602
RESULT	13		
BC010863			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
source			
CDS			

BASE COUNT	613 a	600 c	598 g	622 t
ORIGIN	Query Match 15.5%; Score 49.6; DB 9; Length 2433; Best Local Similarity 48.3%; Pred. No. 0.0018; Matches 139; Conservative 0; Mismatches 149; Indels 0; Gaps 0;			
QY	16	GTATTGCTATGATAACCCAGAGAGATAGAGGTGGAGTTATCAAGTGTTCAGTACTGG	75	
DB	1208	GTGATTGTCATGACACCCCGCTTTGAGAGAGCGGAGGAGAAAGTGTGGCAGTACTGG	1267	
QY	76	CCCGTTTCTCTGAGAGAGCCCTTGGAAATCAAAACATTTTCATGTCCTCTGGAGAACTTT	135	
DB	1268	CCTTTAGAAAAAGACTCTCGGATCCGATTTGGCTTCTCAGAGTGACCAACTTAGGCGTG	1327	
QY	136	CAGTAACCTCAGTATTTTGTCATCCGATATTTCAAATTTGGAAGATCCACAGGAAG	195	
DB	1328	GAGAACATGAATCATTTATAGAAAAACACCGCTAGAAATTCACAAACAGAGGAACGCAG	1387	
QY	196	AGTCACCTCTGTAAACACTTGCAGTTCATCAATATGGCCAGACCATGGCACCTCCCTGCCTCA	255	
DB	1388	AAAGCCAGGTGACCCACTTCCAGTCTTTCAGCTGGCCAGACTATGGTGCTCCCTCTCTCA	1447	
QY	256	GTAGATTTTTTCATCAATATATGTCCTGATGTGAGGAGAGCCACAT	303	
DB	1448	GCAGCTTCCCTCATTGACTTCTTTCAGAGTGTGTGAGAAACAGCAGAGT	1495	
RESULT 14				
HUMPTPSA	3956 bp mRNA PRI 08-JAN-1995			
LOCUS	Human protein-tyrosine phosphatase (PTPase MEG2) mRNA, complete cds.			
DEFINITION	HUMPTPSA			
ACCESSION	M83738			
VERSION	M83738.1 GI:190745			
KEYWORDS	protein-tyrosine phosphatase.			
SOURCE	Homo sapiens placenta cDNA to mRNA.			
ORGANISM	Homo sapiens			
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Gu, M., Warshawsky, I. and Majerus, P. W.			
TITLE	Cloning and expression of a cytosolic megakaryocyte			
JOURNAL	retinaldehyde-binding protein with sequence homology to			
MEDLINE	Proc Natl Acad Sci U S A. 89 (7), 2980-2984 (1992)			
FEATURES	92212952			
source	Location/Qualifiers			
+	1..3956			
+	/organism="Homo sapiens"			
+	/db_xref="taxon:9606"			
+	/cell_line="Meg-01"			
+	/tissue_type="placenta"			
+	509..2290			
+	/gene="PTPase MEG2"			
+	509..2290			
+	/gene="PTPase MEG2"			
+	/codon_start=1			
+	/product="protein-tyrosine phosphatase"			
+	/protein_id="AAA60226.1"			
+	/db_xref="GI:190746"			
+	/translation="MEPATAPRPMAPELTPEEQATKQFLEEINKWTVQYVNSPLSW			
+	NVAKFLMARKDVLRAIELFHSYRETRREGIVKLKPHEEPLRSELSGKFTILNVR			
+	DPFGDLGKVLNLRPHKPSQHVHVLQALFVLDRAVDSFTQRNGLVFIDYMGCSN			
+	ATNEDLSGKVLNLRPHKPSQHVHVLQALFVLDRAVDSFTQRNGLVFIDYMGCSN			
+	TSVTPQHLPRECLPENLGGVYKIDLATWNFQFQPVNGHPDPDEILFLSPALDWD			
+	SYHVPQPHAMTIOELVYVNAKOGTYEYEDIRRENPGVTFHCSMSPGNLEKRNQ			
+	DYPCLDQTVKLTKRSHQTQDYINA SFMDGYKQKNAY IGTOGPLENTYDFWLVWVE			
+	QKVLIVMTVTRFEGRGRGQGVPLEKLSRIDRFELTVNLGVENNNHYKKTITLEI			
+	NTEERQKQVTHFQSWDPGVPSAASLIDFLRVNRQOQSLAVSNMGARSGQGP			
+	PTIVHFCASIGRTGTFCSLDICICLAQLEELGTINLVQTYSRMRTQRAFSIQTPQEQVY			
+	CYKALIEFAEKGMVSSGNLLAVESS"			

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:22:52 ; Search time 171.25 Seconds
(without alignments)
1602.009 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAAATAATTGTAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SID22/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SID22/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SID22/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SID22/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SID22/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SID22/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SID22/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SID22/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SID22/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SID22/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SID22/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SID22/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SID22/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SID22/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SID22/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SID22/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SID22/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SID22/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SID22/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SID22/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SID22/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	100.0	320	20 AAV81747	Rat PTP10 encoding
2	280	87.5	1692	20 AAV81746	Mouse PTP05 isofor
3	280	87.5	1785	20 AAV81744	Mouse PTP05 encodi
4	280	87.5	1896	20 AAV81745	Mouse PTP05 isofor
5	57	17.8	8043	16 AAQ85924	Human protein tyro
6	57	17.8	8119	21 AAC61840	DNA encoding a hum
7	47.2	14.8	5117	16 AAT06027	huDEP-1 cDNA. Hom
c 8	45	14.1	936	22 AAF58252	Oligonucleotide D1
c 9	45	14.1	936	22 AAF58254	Oligonucleotide D1
c 10	45	14.1	936	22 AAF58257	Oligonucleotide D1
c 11	45	14.1	936	22 AAF58259	Oligonucleotide D2

c 12	45	14.1	936	22 AAF58262	Oligonucleotide D2
c 13	45	14.1	938	22 AAF58255	Oligonucleotide D1
c 14	44.6	13.9	4798	22 AAF58418	Human polynucleoti
15	44.4	13.9	936	22 AAF58252	Oligonucleotide D1
16	44.4	13.9	936	22 AAF58254	Oligonucleotide D1
17	44.4	13.9	936	22 AAF58257	Oligonucleotide D1
18	44.4	13.9	936	22 AAF58259	Oligonucleotide D2
19	44.4	13.9	936	22 AAF58262	Oligonucleotide D2
20	44.4	13.9	938	22 AAF58255	Oligonucleotide D1
21	44.2	13.8	2356	20 AAX90696	Human Cytoplasmic
22	44.2	13.8	3058	20 AAX90695	Human Cytoplasmic
23	44.2	13.8	3580	20 AAV81742	Human PTP04 encodi
24	43.4	13.6	3296	15 AAQ55515	PTPase PTP35 genom
25	43.4	13.6	3561	15 AAQ55970	PTPase PTP35 DNA s
26	38.8	12.1	285	17 AAT35309	Rat M2PTP gene cDN
27	36.6	11.4	1191	20 AAX86091	Nucleotide sequenc
28	36.6	11.4	1210	18 AAT95246	Human IA-2/ICA512
29	36.6	11.4	1344	20 AAX86092	Nucleotide sequenc
30	36.6	11.4	1413	20 AAZ32341	Human pancreatic i
31	36.6	11.4	1413	20 AAZ09922	Human islet cell a
32	36.6	11.4	1413	20 AAV63560	Islet cell antibod
33	36.6	11.4	1413	21 AAZ46459	Pancreatic islet c
34	36.6	11.4	3311	20 AAZ32342	Human pancreatic i
35	36.6	11.4	3311	20 AAZ09923	Human islet cell a
36	36.6	11.4	3311	20 AAV63561	Islet cell antibod
37	36.6	11.4	3311	21 AAZ46460	Pancreatic islet c
38	36.6	11.4	3613	20 AAX87627	Human islet cell a
39	36.6	11.4	3984	18 AAT58627	Protein tyrosine p
40	36.6	11.4	3984	20 AAX78463	Human PTPH1 CDNA.
41	36.4	11.4	244	22 AAF58238	Oligonucleotide D1
c 42	34.8	10.9	244	22 AAF58238	Oligonucleotide D1
c 43	34.4	10.8	626	21 AAZ44264	Porcine IFN-gamma
c 44	34.4	10.8	842	19 AAV08186	Porcine interferon
c 45	34.4	10.8	842	20 AAV63230	Porcine interferon

ALIGNMENTS

RESULT 1
AAV81747
ID AAV81747 standard; cDNA; 320 BP.
XX AAV81747;
XX 10-MAR-1999 (first entry)
XX Rat PTP10 encoding cDNA.
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
XX Rattus sp.
XX WO9849317-A2.
XX PD 05-NOV-1998.
XX 27-APR-1998; 98WO-US08439.
XX 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
(SUGEN-) SUGEN INC.

App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
Markby D, Onrust S, Peles E, Plowman GD;

```
XX WPI: 1999-009434/01.
DR P-PSDB; AAW89252.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 149; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes rat PTP10. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 320 BP; 95 A; 64 C; 65 G; 96 T; 0 other;

Query Match 100.0%; Score 320; DB 20; Length 320;
Best Local Similarity 100.0%; Pred. No. 6.4e-89;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATAATTGTAATGTTATTGCTATGATAACCCAGAGAGATAGAAGGTGGAGTTATCAAG 60
DB 1 gaaataaattgtaattgtattgctatgataaaccagagagatagaaggtggagttatcaag 60
QY 61 TGTGTCAGTACTGGCCGCTTCTCTCAAGAGGCTTTGGAATTAACAACATTTTCATGTC 120
DB 61 tgtgtcagtactggccgcttctctcaagagagcctttggaattaaacaacatttcattgctc 120
QY 121 CTCTGTGAGAACTTTCAGATACTCAGTATTGTTGTCATCCGAATATTTCAAATTTGTCGAG 180
DB 121 ctctgtgagaacttccagataactcagattttgtcatccgaatatttc aaattgtggaag 180
QY 181 AAGTCCACAGAAAGACTCTGTGTAAACACTTGCAGTTTCATCAATGCCAGACCAT 240
DB 181 aagtccacagaaagactctgtgtaaacacttgcagtttcacatcaaatggccagaccat 240
QY 241 GGCACCTCCTGCTCAGTGTAGATTTTTCATCAATATGTCGTTATGTGAGGAAGGCCAC 300
DB 241 ggcactcctgctcagtgtagatTTTTCATCAATATGTCGTTATGTGAGGAAGGCCAC 300
QY 301 ATTACAGACCCCTCCCTTGT 320
DB 301 attacagaccctccttgt 320

RESULT 2
AAV81746
ID AAV81746 standard; cDNA; 1692 BP.
XX
XX AAV81746;
XX
XX 10-MAR-1999 (first entry)
XX
XX Mouse PTP05 isoform #2 encoding cDNA.
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
```

```
XX Mus sp.
XX OS
XX FH Key Location/Qualifiers
XX CDS 199..1413
XX FT /*tag= a
XX FT /note= "no stop codon given"
XX
XX WO9849317-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.
XX 11-JUN-1997; 97US-0049477.
XX 11-JUN-1997; 97US-0049756.
XX 18-JUN-1997; 97US-0049914.
XX
XX (SUGE-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B,
XX Markby D, Onrust S, Peles E, Plowman GB;
XX
XX WPI: 1999-009434/01.
XX P-PSDB; AAW89251.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 148-149; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 1692 BP; 507 A; 342 C; 386 G; 457 T; 0 other;

Query Match 87.5%; Score 280; DB 20; Length 1692;
Best Local Similarity 92.2%; Pred. No. 2.4e-76;
Matches 295; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAAATAAATTGTAATGTTATTGCTATGATAACCCAGAGAGATAGAAGGTGGAGTTATCAAG 60
DB 949 gaaataaattgtaattgtattgctatgataaaccagagagatagaaggtggagttatcaag 1008
QY 61 TGTTCAGTTACTGGCCGCTTCTCTGAAGAGCCCTTTGGAATTTCAAACTTTTCATGTC 120
DB 1009 tgttacagttactggccgcttctctgaagagcctttggaattcgaacacttagtgc 1068
QY 121 CTTCCTGGAGAACTTTCAGATACTCAGTATTGTTGTCATCCGAATATTTCAAATTTGTCGAG 180
DB 1069 ttctctggagaccttcagtaactcaatattccacgcttcagtagattgtggaag 1128
QY 181 AAGTCCACAGAAAGACTCTGCTGTAAACACTTGCAGTTTCATCAATGCCAGACCAT 240
DB 181 aagtccacagaaagactctgctgtaaacacttgcagtttcacatcaatgctgagattgtggaag 1128
```

Db 1129 aagtcacagaaagagccaatgtgtaaacacattgcagttcaccaagtgccagaccat 1188
Qy 241 GGCACTCTGCTCAGTAGATTTTTCATCAATATGTCCTTATGTGAGGAGGCCAC 300
|||||
Db 1189 ggcactcgtcctcagcagatgttttccataaaatgcccgttatgtgaggaagaccac 1248
Qy 301 ATTACAGGACCCCTCCTTGT 320
|||||
Db 1249 attacagaccctccttgt 1268

RESULT 3
AAV81744
ID AAV81744 standard; cDNA; 1785 BP.
XX
AC AAV81744;
XX
DT 10-MAR-1999 (first entry)
XX
DE Mouse PTP05 encoding cDNA.
XX
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 199..1479
FT /*tag= a
XX
PN W09849317-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08439.
XX
PR 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
PA (SUGE-) SUGEN INC.
XX
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;
XX
DR WPI; 1999-009434/01.
DR P-PSDB; AAW89249.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
PT
XX
PS Claim 2; Page 147; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in

CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
SQ Sequence 1785 BP; 527 A; 361 G; 408 G; 489 T; 0 other;

Query Match 87.5%; Score 280; DB 20; Length 1785;
Best Local Similarity 92.2%; Pred. No. 2.4e-76;
Matches 295; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 1 GAAATAATTGTAATGTTATTGCTATGATAACCCAGAGATAGAGGTGGAGTTATCAAG 60
|||||
Db 949 gaaataattgtaattgtattgctatgataaccagagagataatgtggagttatcaag 1008
Qy 51 TGTTGCAGTACTGGCCCGTTTCTCTGAAGGAGCCTTTGGAAATTCAAACACTTTTCATGTC 120
|||||
Db 1009 gtttacagttactggccatttctcgaagagaccttggaaattcgaacacttttagtgc 1068
Qy 121 CTTCTGGAGAACTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTTCAAAATGTTGAAG 180
|||||
Db 1069 ttctggagaccttctcatgtaactcaatatttcacccgttcgagttattcagattgtgaag 1128
Qy 181 AAGTCCACAGGAAAGAGTCACTCTGTAAACACACTTGCAAGTTTCATCAATGCGCAGACCAT 240
|||||
Db 1129 aagtcacagaaagagaccatgtgtaaacacattgcagttcaccaagtgccagaccat 1188
Qy 241 GGCACTCTGCTCAGTAGATTTTTCATCAATATGTCCTTATGTGAGGAGGCCAC 300
|||||
Db 1189 ggcactcgtcctcagcagatgttttccataaaatgcccgttatgtgaggaagaccac 1248
Qy 301 ATTACAGGACCCCTCCTTGT 320
|||||
Db 1249 attacagaccctccttgt 1268

RESULT 4
AAV81745
ID AAV81745 standard; cDNA; 1896 BP.
XX
AC AAV81745;
XX
DT 10-MAR-1999 (first entry)
XX
DE Mouse PTP05 isoform #1 encoding cDNA.
XX
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 199..1590
FT /*tag= a
XX
PN W09849317-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08439.
XX
PR 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
PA (SUGE-) SUGEN INC.
XX
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;

XX WPI; 1999-009434/01.
DR P-PSDB; AAW89250.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 147-148; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 1896 BP; 572 A; 372 C; 427 G; 525 T; 0 other;
SQ

Query Match 87.5%; Score 280; DB 20; Length 1896;
Best Local Similarity 92.2%; Pred. No. 2.5e-76;
Matches 295; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 1 GAAATAAATTGTAATGTTATTGCTATGATAACCCAGAGAGATAGAAGTGGAGTTATCAAG 60
Db 1060 gaaaataattgtatgtattgtctatgtatgaacacagagagatagaatgtggagttatacaag 1119
QY 61 TGTTCAGTTACTGGCCGCTTCTCTGAAGGAGCCCTTGGAAATTCACACACTTTTCATGTC 120
Db 1120 tgttacagttactggccattctctgaaggagcccttggaaattcgaacacttagtgctc 1179
QY 121 CTCTCGAGAACCTTCAGATAACTCAGTATTTGTCCATCCGAATATTTCAAATGTGAAG 180
Db 1180 ttcttgagaccttctcatgtaactcaataatttcaccgttcgagtatttcagattgtgaag 1239
QY 181 AAGTCCACAGAAAGAGTCACTCTGTAACACTTGTCAGTTTCATCAAAATGGCCAGACCAT 240
Db 1240 agtccacagaaagagccaatgtgtaaacacttcagttcccaagtggccagaccat 1299
QY 241 GGCACCTCCTGCCTCAGTAGATTTTTCATCAATATGTCCGTTATGTGAGGAAGAGCCAC 300
Db 1300 ggcactcctgcctcagcagattttttcataaataatgctcgtttatgtgaggaagagccac 1359
QY 301 ATTACAGGACCCCTCTTGT 320
Db 1360 attacaggacctctcttgt 1379

RESULT 5
AAQ85924
ID AAQ85924 standard; DNA; 8043 BP.
XX
XX AAQ85924;
XX
XX 05-OCT-1995 (first entry)
XX Human protein tyrosine phosphatase PTP11.
XX
XX Protein tyrosine phosphatase; PTP11; ss.
XX
XX Homo sapiens.
XX

Key Location/Qualifiers
miisc_RNA 45..47
/*tag= a
/note= "in-frame stop codon"
XX
XX W09506735-A.
XX
XX 09-MAR-1995.
XX
XX 01-SEP-1994; 94WO-US09943.
XX
XX 01-SEP-1993; 93US-0115573.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Claesson-welsh L, Gonez LJ, Heldin C, Saras J;
XX WPI; 1995-115450/15.
XX P-PSDB; AAK71498.
XX
XX New protein tyrosine phosphatase(s), PTP11 and GLM-2 - used to
PT develop prods. for use in detection, diagnosis, screening assays
PT or therapy
XX
XX Claim 3; Page 56-68; 92pp; English.
XX
XX A cDNA library was prepd. from human glioma cell line U-343
CC MGA 31L or AG1518 human foreskin fibroblasts. Degenerate primers
CC based on conserved regions of PTPs were prep'd. Sense oligos were
CC AAQ85926 and AAQ85927, and the antisense oligo was AAQ85928. Three
CC clones defined novel PTP sequences. Two of these were named PTP11 and
CC GLM-2. The U-343 MGA 31L cDNA library was screened with labeled probes
CC corresp. to PTP11. The AG1518 human fibroblast cDNA library was also
CC screened. The complete ORF of PTP11 was derived from six
CC overlapping cDNA clones and is given in AAQ85924.
XX
XX Sequence 8043 BP; 2560 A; 1669 C; 1732 G; 2082 T; 0 other;
SQ

Query Match 17.8%; Score 57; DB 16; Length 8043;
Best Local Similarity 50.8%; Pred. No. 1.6e-07;
Matches 163; Conservative 0; Mismatches 155; Indels 3; Gaps 1;
QY 1 GAAATAAATTGTAATGTTATTGCTATGATAACCCAGAGAGATAGAAGTGGAGTTATCAAG 60
Db 6915 gagcaaaaatcccagtgatgcatgactcaagaadtagaagagagaaaaatcaaa 6974
QY 61 TGTTCAGTTACTGGCCGCTTCT 117
Db 6975 tgccagcgtctatggcccaacatctcctaggcaaaacaaatggtcagcaacagactcga 7034
QY 118 GTCTTCTGAGAACCTTTCAGATAAATCTAGTATTTTGTCTATCCGAATATTTCAAATTTGTG 177
Db 7035 ctggctctgtgagaatgcagcagctgaaggccttctgtgtgaggaatgcaccttgaa 7094
QY 176 AAGAAGTCCACAGAAAGAGTCACTCTGTAAAAACACTTGCAGTTTCATCAAAATGGCCAGAC 237
Db 7095 gatattcagaccagagaggtgcgccatattctcatctgaatttcactgcctgcccagac 7154
QY 238 CATGGCAGCTCCTGCCTCAGTAGATTTTTCATCAATATGTCCGTTATGTGAGGAAGAGC 297
Db 7155 catgatcacaccttctcaaccagatgactcttacttttactctctacatgagacacatc 7214
QY 298 CACATTACAGGACCCCTCTCTT 318
Db 7215 cacagatcaggcccaatcatt 7235

RESULT 6
AAC61840
ID AAC61840 standard; DNA; 8119 BP.
XX
XX AAC61840;

Db 3848 gcaatgacatcagaaaattgtttcttcggaatggaccatcagagatttcacagtgaaaaat 3907
QY 181 AAGTCCACAGGAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAAAATGCCAGACCAT 240
Db 3908 atccagacagtgagatcacccctctgagacagttcattccctcctggccagaccac 3967
QY 241 GGACATCCCTGCTCAGTAGATTTTTCATCAAAAT 274
Db 3968 ggtgttcgcagacaccactgacctgctcaact 4001

RESULT 8
AAF58252/C
ID AAF58252 standard; DNA; 936 BP.
XX AC AAF58252;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.

XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAATAATTTGTAATGTTATTTGCTATGATACACAGAGATAGAAGTGGAGTTATCAAGT 61
Db 778 WWWWWW
QY 62 GTTGCAGTTACTGGCCGTTTCTCTGAAGGAGCGCTTTGGAATTCAAACACTTTTCATGTC 121
Db 718 WWWWWW
QY 122 TTCTGGAGAACTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTCAAAATTTGTGAAGA 181
Db 658 WWWWWW

QY 182 AGTCCACAGGAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAAAATGCCAGACCATG 241
Db 598 WWWWWC
QY 242 GCACTCCTGCCTCAGTAGATTTTTCATCAAAATATGTCCGTTATGTGAGGAAGAGCCACA 301
Db 538 WWWWWW
QY 302 TTA 304
Db 478 WWW 476

RESULT 9
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
XX AC AAF58254;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1875.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.

XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAATAATTTGTAATGTTATTTGCTATGATACACAGAGATAGAAGTGGAGTTATCAAGT 61
Db 778 WWWWWW
QY 52 GTTGCAGTTACTGGCCGTTTCTCTGAAGGAGCGCTTTGGAATTCAAACACTTTTCATGTC 121
Db 718 WWWWWW
QY 122 TTCTGGAGAACTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTCAAAATTTGTGAAGA 181


```
Db 658 WWWWWW 599
QY 182 AGTCACAGGAAGAGTCAGTCTGTAAACACTTCAGTTCATCAAAATGGCCAGACCACATG 241
Db 598 WWWWWW 539
QY 242 GCATCTCCCTCAGTACAGTATTTTCATCAAAATATGTCGTTATCTGAGGAGAGCCACACA 301
Db 538 WWWWWW 479
QY 302 TTA 304
Db 478 WWW 476

RESULT 10
AAF58257/C
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAAATAATTGTAATCTTATGCTATGATAACACAGAGATAGAGTGGAGTTATCAAGT 61
Db 778 WWWWWW 719
QY 62 GTTCAGTTACTGGCCCGTTTCTCTGAAGAGCCCTTTGGAATTCACACATTTTCATGTCC 121
Db 718 WWWWWW 659
```

```
QY 122 TTCTCGGAAACTTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTTCAAATGTGAAGA 181
Db 638 WWWWWW 599
QY 182 AGTCACAGGAAGAGTCACTCTGTAAACACTTCAGTTCATCAAAATGGCCAGACCACATG 241
Db 598 WWWWWW 539
QY 242 GCATCTCCCTCAGTACAGTATTTTCATCAAAATATGTCGTTATCTGAGGAGAGCCACACA 301
Db 538 WWWWWW 479
QY 302 TTA 304
Db 478 WWW 476

RESULT 11
ZAF58259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAAATAATTGTAATCTTATGCTATGATAACACAGAGATAGAGTGGAGTTATCAAGT 61
Db 778 WWWWWW 719
QY 62 GTTCAGTTACTGGCCCGTTTCTCTGAAGAGCCCTTTGGAATTCACACATTTTCATGTCC 121
Db 718 WWWWWW 659
```


[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM nucleic - nucleic search, using sw model
Run on: March 30, 2002, 06:07:33 ; Search time 175.91 Seconds
(without alignments)
1559.571 Million cell updates/sec

Title: US-09-095-478A-4
Perfect score: 320
Sequence: 1 GAAATAATGTAATGTTAT.....ATTACAGGACCCCTCCTTGT 320

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	100.0	320	AAV81747	Rat PTP10 encoding
2	50	15.6	1692	AAV81746	Mouse PTP05 isofor
3	50	15.6	1785	AAV81744	Mouse PTP05 encodi
4	50	15.6	1896	AAV81745	Mouse PTP05 isofor
5	33	10.3	33	AAV81754	Human PTP04 RACE p
6	33	10.3	33	AAV81755	Human PTP04 RACE p
7	21	6.6	477	AAH94038	Human foetal cDNA,
8	21	6.6	1184	AAH94456	Human foetal cDNA,
9	18	5.6	192	AAT21420	Human gene signatu
10	18	5.6	304	AAA44583	Human secreted exp
11	18	5.6	332	AAH50777	Human tumour assoc

12	18	5.6	1398	20	AAX02011	Human HDRC DNA. H
13	17	5.3	182	22	AAI54116	Probe #22802 used
14	17	5.3	528	22	AAF68447	Human lung tumour
15	17	5.3	538	22	AAI1051	Probe #9737 used t
16	17	5.3	542	22	AAH11751	Human CDNA clone (
17	17	5.3	587	22	AAH09216	Human CDNA clone (
18	17	5.3	696	21	AAF13810	Aspergillus oryzae
19	17	5.3	718	22	AAH98609	Human EST-derived
20	17	5.3	749	22	AAH07437	Human CDNA clone (
21	17	5.3	785	11	AAQ04721	Human arginase cDN
22	17	5.3	878	21	AAF18118	Human arginase cDN
23	17	5.3	931	21	AAZ98037	Lung cancer associ
24	17	5.3	931	22	AAH11650	Human secreted pro
25	17	5.3	966	11	AAQ04714	Human secreted pro
26	17	5.3	969	19	AAV37940	Human arginase enc
27	17	5.3	972	19	AAV37941	Rat arginase encod
28	17	5.3	1365	11	AAQ04715	Human arginase cDN
29	17	5.3	1437	22	AAH57465	Human liver cell s
30	17	5.3	1563	21	AAH43257	Arabidopsis thalia
31	17	5.3	1736	21	AAH69104	Human secreted pro
32	17	5.3	1772	22	AAH16504	Human CDNA sequenc
33	17	5.3	1848	21	AAH50527	Arabidopsis thalia
34	17	5.3	2922	22	AAH18353	Human CDNA sequenc
35	17	5.3	3299	22	AAH18556	Human CDNA sequenc
36	17	5.3	5558	19	AAV52236	Streptococcus pneu
37	17	5.3	138169	21	AAH34791	Human adenosine re
38	17	5.3	141589	21	AAF20913	Human ELAM-1 polyn
39	17	5.3	141589	21	AAF21127	Human low adenosin
40	17	5.3	141589	21	AAF21152	Human low adenosin
41	17	5.3	141589	21	AAA35005	Human adenosine re
42	17	5.3	141589	21	AAA35030	Human adenosine re
43	17	5.3	143068	21	AAF21105	Human low adenosin
44	17	5.3	143068	21	AAF21272	Human low adenosin
45	17	5.3	143068	21	AAA34983	Human adenosine re

ALIGNMENTS

RESULT : 1
AAV81747
ID AAV81747 standard; cDNA; 320 BP.
AC AAV81747;
XX ..
DT 1C-MAR-1999 (first entry)
DE Rat PTP10 encoding cDNA.
XX ..
DE PTP04; PTP05; PTP10; SAD; ALP; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
XX ..
OS Rattus sp.
XX ..
PN WO9849317-A2.
XX ..
XX PD 05-NOV-1998.
XX ..
PF 27-APR-1998; 98WO-US08439.
XX ..
PR 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX ..
(SUGEN-) SUGEN INC.
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plozman GD;

AAV81754
ID AAV81754 standard; DNA; 33 BP.

XX AC AAV81754;

XX DT 10-MAR-1999 (first entry)

XX DE Human PTP04 RACE primer 6599.

XX KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
neurodegenerative disease; neuronal survival; Alzheimer's disease;
Parkinson's disease; Huntington's disease; PCR primer; ss.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9849317-A2.

XX PD 05-NOV-1998.

XX PF 27-APR-1998; 98WO-US08439.

XX PR 23-OCT-1997; 97US-0063595.

XX PR 28-APR-1997; 97US-0044428.

XX PR 20-MAY-1997; 97US-0047222.

XX PR 11-JUN-1997; 97US-0049477.

XX PR 11-JUN-1997; 97US-0049756.

XX PR 18-JUN-1997; 97US-0049914.

XX PA (SUGEN-) SUGEN INC.

XX PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

XX PI Markby D, Onrust S, Peles E, Plowman GD;

XX XX WPI; 1999-009434/01.

XX XX New nucleic acid encoding specific protein tyrosine phosphatases -

PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease

XX XX Example 1; Page 74; 193pp; English.

XX CC The present invention describes isolated, enriched or purified nucleic
acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
above proteins, other than ALK-7, are protein tyrosine phosphatases
(PTPs) and are used to identify substances that modulate their activity
(i.e. agonists and antagonists, including NBP) in vivo or in vitro.
These substances are used to treat or prevent diseases associated with
abnormal signal transduction pathways that involve the proteins,
particularly cancer (e.g. leukaemia and lymphoma), while modulators of
ALK-7 (which is a type I receptor serine/threonine kinase) are used to
promote neuronal survival, particularly for treating Alzheimer's,
Parkinson's or Huntington's diseases. Nucleic acid fragments of the
polynucleotides encoding the proteins can be used as probes to identify
and clone related sequences; to detect protein-encoded RNA; to generate
transgenic animals and in gene therapy (optionally after mutation). Ab
are used to determine the proteins. The present sequence represents a
RACE primer for human PTP04.

XX SQ Sequence 33 BP; 10 A; 8 C; 9 G; 6 T; 0 other;

Query Match 10.3%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CCGTATGTGAGGAGAGCCACATTACAGGACC 311

Db 1 ccgtatgtgagagagccacattacaggacc 33

RESULT 6

AAV81755/C

XX ID AAV81755 standard; DNA; 33 BP.

XX AC AAV81755;

XX DT 10-MAR-1999 (first entry)

XX DE Human PTP04 RACE primer 6600.

XX KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
neurodegenerative disease; neuronal survival; Alzheimer's disease;
Parkinson's disease; Huntington's disease; PCR primer; ss.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9849317-A2.

XX PD 05-NOV-1998.

XX PF 27-APR-1998; 98WO-US08439.

XX PR 23-OCT-1997; 97US-0063595.

XX PR 28-APR-1997; 97US-0044428.

XX PR 20-MAY-1997; 97US-0047222.

XX PR 11-JUN-1997; 97US-0049477.

XX PR 11-JUN-1997; 97US-0049756.

XX PR 18-JUN-1997; 97US-0049914.

XX PA (SUGEN-) SUGEN INC.

XX PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

XX PI Markby D, Onrust S, Peles E, Plowman GD;

XX XX WPI; 1999-009434/01.

XX XX New nucleic acid encoding specific protein tyrosine phosphatases -

PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease

XX XX Example 1; Page 74; 193pp; English.

XX CC The present invention describes isolated, enriched or purified nucleic
acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
above proteins, other than ALK-7, are protein tyrosine phosphatases
(PTPs) and are used to identify substances that modulate their activity
(i.e. agonists and antagonists, including NBP) in vivo or in vitro.
These substances are used to treat or prevent diseases associated with
abnormal signal transduction pathways that involve the proteins,
particularly cancer (e.g. leukaemia and lymphoma), while modulators of
ALK-7 (which is a type I receptor serine/threonine kinase) are used to
promote neuronal survival, particularly for treating Alzheimer's,
Parkinson's or Huntington's diseases. Nucleic acid fragments of the
polynucleotides encoding the proteins can be used as probes to identify
and clone related sequences; to detect protein-encoded RNA; to generate
transgenic animals and in gene therapy (optionally after mutation). Ab
are used to determine the proteins. The present sequence represents a
RACE primer for human PTP04.

XX SQ Sequence 33 BP; 6 A; 9 C; 8 G; 10 T; 0 other;

Query Match 10.3%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CCGTATGTGAGGAGAGCCACATTACAGGACC 311

Db 33 CCGTATGTGAGGAGAGCCACATTACAGGACC 1

RESULT 7

AAH94038
ID AAH94038 standard; cDNA: 477 BP.
XX
AC AAH94038;
XX
DT 05-OCT-2001 (first entry)
XX
DE Human foetal cDNA, SEQ ID NO: 567.
XX
KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW neotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
PN WO200155339-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02723.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 15-SEP-2000; 2000US-0663870.
PR 06-NOV-2000; 2000US-0707351.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX
DR WPT: 2001-465571/50.
DR P-PSDB; AAM06781.
XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -
XX
PS Claim 1; Page 385; 715pp; English.
XX
CC The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence was assembled using an expressed sequence tag (EST) found
CC to be expressed in human foetal tissue cDNA libraries as the seed.
XX
SQ Sequence 477 BP; 128 A; 90 C; 78 G; 180 T; 1 other;

Query Match 6.6%; Score 21; DB 22; Length 477;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 TTCAAAACACTTTCATGTCCTT 123
Db 281 ttcaaacactttcatgtcctt 301

RESULT 8
AAH94456
ID AAH94456 standard; cDNA: 1184 BP.
XX
AC AAH94456;
XX
DT 05-OCT-2001 (first entry)
XX
DE Human foetal cDNA, SEQ ID NO: 1143.
XX
KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW neotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
PN WO200155339-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02723.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 15-SEP-2000; 2000US-0663870.
PR 06-NOV-2000; 2000US-0707351.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX
DR WPT: 2001-465571/50.
DR P-PSDB; AAM06781.
XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -
XX
PS Claim 1; Page 385; 715pp; English.
XX
CC The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence was assembled using an expressed sequence tag (EST) found
CC to be expressed in human foetal tissue cDNA libraries as the seed.
XX
SQ Sequence 477 BP; 128 A; 90 C; 78 G; 180 T; 1 other;

Query Match 6.6%; Score 21; DB 22; Length 477;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 TTCAAAACACTTTCATGTCCTT 123
Db 281 ttcaaacactttcatgtcctt 301

RESULT 9
AAH94456
ID AAH94456 standard; cDNA: 1184 BP.
XX
AC AAH94456;
XX
DT 02-AUG-1996 (first entry)
XX
DE Human gene signature HUMGS02790.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX

```
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
PI WPI; 1995-206931/27.
XX
DR
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 882; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 192 BP; 67 A; 25 C; 21 G; 77 T; 2 other;

Query Match 5.6%; Score 18; DB 16; Length 192;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATAATATTCATGCTT 18
Db |||||
111 gaaataatgtaattgtt 128

RESULT 10
AAA44583
ID AAA44583 standard; cDNA; 304 BP.
XX
AC AAA44583;
XX
XX 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:1158.
XX
XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiaschmatic; vulnary; antiparkinsonian;
XX antiulcer; osteoprotective; neuroprotective; nootropic; antipsoriatic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX
```

```
OS Homo sapiens.
XX
XX WC200021991-A1.
XX
XX 20-APR-2000.
XX
XX 15-OCT-1999; 99WO-US24206.
XX
XX 15-OCT-1998; 98US-0104436.
XX
XX (GEM ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Bowman MR;
XX WPI; 2000-317938/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (SESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 500; 803pp; English.
XX
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
XX sequence tags (SESTs), isolated from human, mouse, chicken and rat
XX tissue sources. The SESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemokinetic; proliferative; immunomodulatory; haematopoietic;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnary; antiulcer; osteoprotective; neuroprotective;
XX nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The SESTs can be used for gene
XX therapy and in vaccines. The SESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
XX (Alzheimer's, Parkinson's, Huntington's, inflammatory disorders (Crohn's
XX disease), tumours, bacterial, fungal or viral infections, depression and
XX psoriasis. AAA45926 to AAA45931 represent linker variants which are given
XX in the exemplification of the present invention.
XX
SQ Sequence 304 BP; 81 A; 67 C; 75 G; 81 T; 0 other;

Query Match 5.6%; Score 18; DB 21; Length 304;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 CAACACTTTCATGCTCT 122
Db |||||
171 caaacacttctcatgtct 188

RESULT 11
AAH50777
ID AAH50777 standard; cDNA; 332 BP.
XX
AC AAH50777;
XX
XX 23-AUG-2001 (first entry)
XX
XX Human tumour associated cDNA #106.
XX
XX Human; cancer specific gene expression; gene therapy;
XX age related differential expression; ss.
XX
OS Homo sapiens.
```

```

XX PN WO200136685-A2.
XX DR
XX PD
XX PF
XX PR 17-NOV-2000; 2000WO-US31809.
XX PR 17-NOV-1999; 99US-0166056.
XX PR 17-NOV-1999; 99US-0166106.
XX PA (NYXI-) NYXIS NEURO THERAPIES INC.
XX PI Kroes RA, Moskal JR, Yamamoto H;
XX DR WPI; 2001-355647/37.
XX PT Novel nucleic acid molecules differentially expressed in brain cancers,
XX PT useful for ascertaining propensity of cell for malignant phenotype or
XX PT ascertaining suitability of anti-neoplastic drug candidate -
XX PS Claim 28; Page 52; 82pp; English.
XX CC The present invention provides the sequences of 184 cDNA fragments which
XX CC are differentially expressed in cancer cell depending on the age of the
XX CC patient. They can be used to diagnose and identify treatments for
XX CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,
XX CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The
XX CC present sequence is a cancer-associated cDNA of the invention.
XX SQ Sequence 332 BP; 95 A; 45 C; 72 G; 120 T; 0 other;

Query Match 5.6%; Score 18; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TCAGATTGTGAGAGTC 185
Db 79 tcaaatgtgaagaagtc 96

RESULT 12
AAX02011
ID AAX02011 standard; DNA; 1398 BP.
AC AAX02011;
XX 21-APR-1999 (first entry)
DE Human HDRG DNA.
XX Human developmentally regulated GTP-binding protein; HDRG; cancer;
KW diagnosis; treatment; prevention; immune disorder; neoplastic disorder;
KW smooth muscle disorder; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 64..1167
FT /*tag= a
FT /product= "HDRG"
XX US5871971-A.
PN 16-FEB-1999.
PD 17-JUL-1997; 97US-0896320.
XX 17-JUL-1997; 97US-0896320.
PR (INCY-) INCYTE PHARM INC.
XX Corley NC, Hillman JL, Lal P, Shah P;

```

```

XX DR WPI; 1999-166645/14.
XX DR P-ESDB; AAW92455.
XX PT New polynucleotides encoding human developmentally regulated
XX PT GTP-binding protein (HDRG) - useful for diagnosing, preventing and
XX PT treating immune, abnormal vesicle trafficking, smooth muscle and
XX PT neoplastic disorders
XX PS Claim 1; Fig 1A-D; 32pp; English.
XX CC This sequence encodes a human developmentally regulated GTP-binding
XX CC protein (HDRG). The HDRG protein can be used as a probe to detect a
XX CC polynucleotide which encodes HDRG. The polynucleotide sequences encoding
XX CC HDRG (or complements) may be used for the diagnosis of conditions or
XX CC disorders associated with expression of HDRG and treatment/prevention
XX CC of immune, abnormal vesicle trafficking, smooth muscle and neoplastic
XX CC disorders e.g. AIDS, Addison's disease, atherosclerosis, bronchitis,
XX CC ulcerative colitis, diabetes mellitus, emphysema, gout, Graves' disease,
XX CC osteoporosis, rheumatoid arthritis, Sjogren's syndrome and
XX CC complications of cancer. The polynucleotides may also be used in Southern
XX CC or Northern analysis, dot blot, or other membrane based technologies, in
XX CC PCR technologies or in dipstick, pin, ELISA assays or microarrays
XX CC utilizing fluids or tissues from patient biopsies to detect altered
XX CC HDRG expression.
XX SQ Sequence 1398 BP; 394 A; 302 C; 354 G; 348 T; 0 other;

Query Match 5.6%; Score 18; DB 20; Length 1398;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 TTCAAATTGTGAGAGT 184
Db 1148 ttcaaatgtgaagaagt 1165

RESULT 13
AAX02011
ID AAX02011 standard; DNA; 182 BP.
AC AAX02011;
XX 17-OCT-2001 (first entry)
DE Probe #22802 used to measure gene expression in human placenta sample.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX Homo sapiens.
XX WQ200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00663.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX

```

PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 22802; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 182 BP; 50 A; 36 C; 58 G; 38 T; 0 other;

Query Match 5.3%; Score 17; DB 22; Length 182;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 TTTCAAATTCGTGAAGAA 182
 |||||
 Db 60 ttccaattgtgaagaa 76

RESULT 14
 AAF68447
 ID AAF68447 standard; cDNA; 528 BP.
 XX
 AC AAF68447;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Human lung tumour protein related nucleotide sequence SEQ ID NO:382.
 XX
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 KW cytostatic; antisense inhibition; ss.
 XX
 OS Homo sapiens.

XX WO200100828-A2.
 XX 04-JAN-2001.
 XX 30-JUN-2000; 2000WO-US18061.
 XX 30-JUN-1999; 99US-0346492.
 PR 15-OCT-1999; 99US-0419356.
 PR 17-DEC-1999; 99US-0466867.
 PR 30-DEC-1999; 99US-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-0533077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MW, Mannion J;
 XX WPI; 2001-071488/08.
 XX
 XX Lung tumor-associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -
 PT
 XX
 XX Claim 4; Page 285; 436pp; English.
 XX
 CC The present invention describes immunogenic portions of lung tumour-
 CC associated proteins (I) and the nucleic acids (NAS) that encode them.
 CC (I) have cytostatic activity and can be used in gene therapy, antisense
 CC inhibition and in vaccines. The NAS and the lung tumour-associated
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with their inappropriate expression,

CC especially lung cancers. For example, the NAS may be administered to
 CC treat diseases by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of the protein by expressing inactive proteins
 CC or to supplement the patients own production of (I). Additionally, the
 CC NAS may be used to produce the lung-tumour associated protein, according
 CC to standard recombinant DNA methodology. Conversely, antisense NA
 CC molecules may be administered to down regulate protein expression by
 CC binding with the cells own genes and preventing their expression. The NA
 CC and complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar NA sequences in
 CC samples, and hence which patients may be in need of treatment for lung
 CC cancer. The (I) may be used as antigens in the production of antibodies
 CC and in assays to identify modulators (agonists and antagonists) of the
 CC expression and activity of the protein. AAF68083 to AAF68878 and
 CC AAF76848 to AAF76878 represent human lung tumour protein related
 CC nucleotide and protein sequences which are used in the exemplification
 CC of the present invention.

SQ Sequence 528 BP; 112 A; 141 C; 143 G; 130 T; 2 other;

Query Match 5.3%; Score 17; DB 22; Length 528;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 GCCTCAGTAGATTTT 266
 |||||
 Db 330 gcctcagtagatittt 346

RESULT 15
 AA141051
 ID AA141051 standard; DNA; 538 BP.
 XX
 AC AA141051;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #9737 used to measure gene expression in human placenta sample.
 XX Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.

XX WC200157272-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00663.
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488897/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 PT
 XX
 XX Claim 25; SEQ ID No 9737; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX
SQ Sequence 538 BP; 159 A; 104 C; 123 G; 152 T; 0 other;

Query Match 5.3%; Score 17; DB 22; Length 538:

Best Local Similarity 100.0%; Pred. No. 34;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 TTTCAAATTGTGAGAA 182

Db 378 ttccaattgtgaagaa 394

Search completed: March 30, 2002, 07:13:51
Job time: 3978 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:13:48 ; Search time 119.38 Seconds
(without alignments)
149.483 Million cell updates/sec

Title: US-09-095-478a-8
Perfect score: 673
Sequence: 1 DFGMMWNNCNIAMITRE.....VRKSHITGLLVHCTAGVGR 122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL.17.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	566	84.1	426	11	O55082	O55082 mus musculus
2	529	78.6	398	4	Q9Y406	Q9Y406 homo sapien
3	284.5	42.3	2484	6	Q28006	Q28006 bos taurus
4	274.5	40.8	2460	11	Q64512	Q64512 mus musculus
5	261	38.8	1156	5	Q21214	Q21214 caenorhabdi
6	260.5	38.7	121	4	Q9UDA8	Q9UDA8 homo sapien
7	255.5	38.0	1216	13	Q9IBD8	Q9IBD8 cyprinus ca
8	253.5	37.7	1245	13	Q9IBF0	Q9IBF0 fugu rubrip
9	253.5	37.7	1246	13	Q9IBF1	Q9IBF1 fugu rubrip
10	243.5	36.2	578	5	Q9W324	Q9W324 drosophila
11	240.5	35.7	597	13	Q9IB71	Q9IB71 xenopus lae
12	240	35.7	579	11	Q9JJ07	Q9JJ07 mus musculus
13	239	35.5	1337	4	Q15255	Q15255 homo sapien
14	238	35.4	832	13	Q90947	Q90947 gallus gall
15	238	35.4	1648	4	Q9UDR6	Q9UDR6 homo sapien
16	237	35.2	468	13	Q9IBA0	Q9IBA0 potamotrygo
17	237	35.2	469	13	Q9NLO8	Q9NLO8 eptatretus
18	237	35.2	615	13	Q9IAI8	Q9IAI8 xenopus lae
19	237	35.2	1399	4	O75870	O75870 homo sapien

20	237	35.2	1499	13	Q90815	Q90815 gallus gall
21	237	35.2	1502	4	Q90M81	Q90M81 homo sapien
22	237	35.2	1948	4	Q13332	Q13332 mus musculus
23	235.5	35.0	1291	11	Q61812	Q61812 mus musculus
24	235	34.9	468	13	Q9NLO6	Q9NLO6 eptatretus
25	233.5	34.7	694	13	Q91870	Q91870 xenopus lae
26	233	34.6	1216	11	Q62884	Q62884 rattus norv
27	232.5	34.5	811	4	Q9UI21	Q9UI21 homo sapien
28	232.5	34.5	813	4	Q9UI22	Q9UI22 homo sapien
29	232.5	34.5	823	4	Q9Y4V5	Q9Y4V5 homo sapien
30	232.5	34.5	1436	4	Q92850	Q92850 homo sapien
31	232.5	34.5	1436	4	O00197	O00197 homo sapien
32	232.5	34.5	1439	4	P78399	P78399 homo sapien
33	232	34.5	361	11	Q61373	Q61373 mus musculus
34	232	34.5	468	13	Q9IBA2	Q9IBA2 potamotrygo
35	232	34.5	1896	13	Q9IAJ1	Q9IAJ1 xenopus lae
36	231	34.3	1501	11	Q9QW00	Q9QW00 rattus sp.
37	231	34.3	1863	11	Q64605	Q64605 rattus norv
38	231	34.3	1904	11	Q64699	Q64699 mus musculus
39	230.5	34.2	1114	4	Q9H0Y6	Q9H0Y6 homo sapien
40	230.5	34.2	1143	4	Q15614	Q15614 homo sapien
41	230	34.2	460	11	Q62917	Q62917 rattus norv
42	230	34.2	582	11	Q64696	Q64696 mus musculus
43	230	34.2	1406	13	Q9W6V5	Q9W6V5 gallus gall
44	230	34.2	1887	11	Q9QW67	Q9QW67 rattus sp.
45	230	34.2	1898	11	Q64604	Q64604 r protein-t

ALIGNMENTS

RESULT 1
Q55082
ID O55082 PRELIMINARY; PRT; 426 AA.
AC O55082;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 20 (EC 3.1.3.48)
DE (PHOSPHOTYROSINE PHOSPHATASE) (PTPASE).
GN PTPN20.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
GX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=98070510; PubMed=9407093;
RA Ohsugi M., Kuramochi S., Matsuda S., Yamamoto T.;
RT "Molecular cloning and characterization of a novel cytoplasmic
RT n-protein-tyrosine phosphatase that is specifically expressed in
RT spermatocytes."
RL J. Biol. Chem. 272:33092-33099(1997).
CC - CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -> PROTEIN
CC - TYROSINE + PHOSPHATE.
DR EMBL; D64141; BAA23761.1; -
DR HSSP; P28827; IRPM.
DR MGD; MGI:1196295; Ptpn20.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPHTASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
SQ SEQUENCE 426 AA; 49118 MW; 2B35FBI3379502F4 CRC64;

Query/Match 84.1%; Score 566; DB 11; Length 426;
Best Local Similarity 83.6%; Pred. No. 1.le-55;

Der 26

```

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
RA Vega Q.C., Walton K.M., Dixon J.E.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN
CC -!- TYROSINE + PHOSPHATE.
CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC RADIXIN, AND TALIN.
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
DR EMBL: U20807; AAA73516.1; -.
DR HSP; Q12923; 3PZ.
DR InterPro: IPR000299; Band_4.1.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000387; Tyr_phosphatase.
DR InterPro: IPR000242; Tyr_prot_phptase.
DR Pfam: PF00373; Band_4.1; 1.
DR Pfam: PF00595; PDZ; 5.
DR Pfam: PF00102; Y-phosphatase; 1.
DR PRINTS: PR00935; BAND41.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00295; B41; 1.
DR SMART: SM00228; PDZ; 5.
DR SMART: SM00194; PTPC; 1.
DR PROSITE: PS00660; BAND_4.1; FALSE_NEG.
DR PROSITE: PS00661; BAND_4.1.2; FALSE_NEG.
DR PROSITE: PS50057; BAND_4.1_3; 1.
DR PROSITE: PS50106; PDZ; 5.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Structural protein; Cytoskeleton; Hydrolase; Coiled coil.
FT DOMAIN 600 800
FT DMAIN 1341 1344 POLY-SER.
FT DMAIN 2236 2484 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 2407 2407 BY SIMILARITY..
FT DMAIN 467 504 COILED COIL (POTENTIAL).
FT DMAIN 1761 1793 COILED COIL (POTENTIAL).
SQ SEQUENCE 2484 AA; 276382 MW; 45A92F0D4F1ED13D CRC64;

Query Match 42.3%; Score 284.5; DB 6; Length 2484;
Best Local Similarity 42.3%; Pred. No. 4.6e-23;
Matches 52; Conservative 21; Mismatches 49; Indels 1; Gaps

QY 1 DFWMGMWNNCNVNIAMITREGGVIKCSYWPVSL-KEPLEFKHFHVLLLENFOITOYFY 59
Db 2291 DFQMIMEONSSVIMMTQVEGEIKICQRTWPNVLGKSTWVSNRLIALRVQOLKGKV 2350
QY 60 IRIFIVKKTSKSHSVKHLOFIKPDPHGCTPASVDFFTKYVRYVRKSHITGTGLLVHCTAG 119
Db 2351 VRAMTLEDIQTEVHRVSHLNFTAPDDHTDSQPDDLTLTFYSWMRHVSRSPIITHCSAG 2410
QY 120 VGR 122
Db 2411 IGR 2413

RESULT 4
ID Q64512 PRELIMINARY; PRT; 2460 AA.
AC Q64512; Q62135; Q61494; Q64499;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR-TYPE, 13 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE RIP) (PHOSPHOPROTEIN PHOSPHATASE)
DE (PROTEIN-TYROSINE-PHOSPHATASE) (PHOSPHOTYROSINE PHOSPHATASE) (PTPASE)

```


RP SEQUENCE FROM N.A.
RX MEDLINE-94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans*;
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; Z69664; CAA93513.1; --
DR HSP; O66124; 25HP
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003600; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000531; TonB_boxC.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG-like; 1.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
SQ SEQUENCE 1156 AA; 130353 MW; 688AA2A085980691 CRC64;

Query Match 38.8%; Score 261; DB 5; Length 1156;
Best Local Similarity 40.3%; Pred. No. 8.6e-21;
Matches 50; Conservative 16; Mismatches 56; Indels 2; Gaps 1;
QY 1 DFGGMWNNCNCVIAITREIEGGVVKCCSWPVSLKEPLEFKHFHVLLENFQITQYFVI 60
DB 742 DFWSMINQERSNIIVICITNVEDCKRKCDQYWPQSDSPQTFGNYQVTLYSESTNAHF 801
QY 61 RI--FQIVKSTGKSHVKHLOFTKWDHGTTPASVDFFIKYRVYRKSHITGPLLHCTA 118
DB 802 RILDKIAKAVPAVERKVKHOLHFGWPDHGVPSVFPFLLSFVHTSDIHSTGTPVWHCSA 861
QY 119 GVGR 122
DB 862 GVGR 865
RESULT 6
ID Q9UDA8 PRELIMINARY; PRT; 121 AA.
AC Q9UDA8
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DE 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RA MEDLINE-93247293; PubMed=8483328;
RA Honda H., Shibuya M., Chiba S., Yazaki Y., Hirai H.;
RT *Identification of novel protein-tyrosine phosphatases in a human
RL Leukemia cell line, F-36P*;
DR HSEp; P18052; LYFO.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
SQ SEQUENCE 121 AA; 14205 MW; 4F9665F6F9348055 CRC64;
Query Match 38.7%; Score 260.5; DB 4; Length 121;
Best Local Similarity 39.7%; Pred. No. 7.5e-22;
Matches 48; Conservative 20; Mismatches 52; Indels 1; Gaps 1;
QY 3 WGMWNNCNCVIAITREIEGGVVKCCSWPVSLKEPLEFKHFHVLLENFQITQYFVIR 61
DB 1 WRWMEOKSTVIAMWTOEVEGEKIKQRYWPNILKTTWVSNRLRLALVRMQLKGFVVR 60
QY 62 IFQIVKSTGKSHVKHLOFTKWDHGTTPASVDFFIKYRVYRKSHITGPLLHCTAGVG 121
DB 61 AMTLEDIQTRVVRHSHLNFATWPDHDPSPQDDLLTETISYMRHHRSGPIITHCSAGIG 120
QY 122 R 122
DB 121 R 121
RESULT 7
ID Q9IBD8 PRELIMINARY; PRT; 1216 AA.
AC Q9IBD8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DE 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CD45.
OC Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
CX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RA Fujiki K., Nakao M., Shin D., Yano T.;
RT *Molecular cloning of carp (Cyprinus carpio) leukocyte cell-derived
RT chemotaxin 2, glia maturation factor beta, CD45, and lysozyme C by use
RT of suppression subtractive hybridization*;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB031424; BAA92179.1;
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR000387; Tyr_phosphatase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00194; PTPc; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
SQ SEQUENCE 1216 AA; 138251 MW; 9D2E45F41721D2CF CRC64;
Query Match 38.0%; Score 255.5; DB 13; Length 1216;
Best Local Similarity 40.8%; Pred. No. 3.8e-20;
Matches 51; Conservative 19; Mismatches 52; Indels 3; Gaps 2;
QY 1 DFGGMWNNCNCVIAITREIEGGVVKCCSWPVSLKEPLEFKHFHVLLENFQITQYFVI 60

Db	642	DFWRMVQKSSIIIVWVTRCEEENKTKCAQYWPSLDRERAEI	FEFVVKIRSEBHC	PDYVI	700
Oy	61	R-IFQIVKKSQSHSVKHLQFIKWDPDHPGTPASVDFFIKVVYVR--	KSHITG	GPLLVHCT	117
Db	702	RHLILNNKREKGSEREVTHIQFI	SWPDHGVGPGDPSLLLLKLR	RRVNSFKNFFSG	PVVHCS 761
Oy	118	AGVGR	122		
Db	762	AGVGR	766		
RESULT	8				
Q918F0					
ID	Q918F0	PRELIMINARY;	PRT;	1245	AA.
AC	Q918F0;				
DT	01-OCT-2000	(TReMBLrel. 15, Created)			
DT	01-OCT-2000	(TReMBLrel. 15, Last sequence update)			
DT	01-JUN-2001	(TReMBLrel. 17, Last annotation update)			
DE	CD45	(EC 3.1.3.48).			
GN	PTPC.				
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
OC	Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;				
OC	Tetraodontidae; Takifugu.				
OX	NCBI_TaxID=31033;				
RP	[1]				
RN	SEQUENCE FROM N.A.				
RC	TISSUE=SPLEEN;				
RA	Diaz del Pozo E.M., Beverley P.C., Timon M.;				
RT	Genomic structure and sequence of the leukocyte common antigen (CD45) from the pufferfish Fugu rubripes and comparison with its mammalian homologue."				
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AJ243430; CAB96212.1; "				
DR	InterPro: IPR003961; FN_III.				
DR	InterPro: IPR000387; Tyr_phosphatase.				
DR	InterPro: IPR000242; Tyr_prot_phptase.				
DR	Pfam: PF00041; fn3; 1.				
DR	Pfam: PF00102; Y_phosphatase; 2.				
DR	PRINTS: PR00700; PRTYPHPHTASE.				
DR	SMART: SM00060; FN3; 1.				
DR	SMART: SM00194; PTPC; 2.				
DR	PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.				
DR	PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.				
DR	PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.				
KW	Hydrolase.				
QY	SEQUENCE	1245	AA;	141324	MW; 6CB71LEFF5797555
					CRC64;
Query Match		37.7%;	Score	253.5;	DB 13; Length 1245;
Best Local Similarity		39.2%;	Pred.	No. 6.5e-20;	
Matches	49;	Conservative	21;	Mismatches	52;
				Indels	3;
				Gaps	
Oy	1	DFWGMWENNVCNVIAMITREIEGGVIKCSYMPVSLKEP	LFKHFHVLLENFQITQY	FEVI	60
Db	700	NFRMIWEQOTSIIIVWVTRCEEENKTKCAQYWPSLDRERAEI	FEFVVKIRSEBHC	PDYVI	759
Oy	61	RIFQIV-KKSTGKSHSVKHLQFIKWDPDHPGTPASVDFFIKVVYVR--	KSHITG	GPLLVHCT	117
Db	760	RHLILNNKREKGSEREVTHIQFI	SWPDHGVGPGDPSLLLLKLR	RRVNSFKNFFSG	PVVHCS 819
Oy	118	AGVGR	122		
Db	820	AGVGR	824		
RESULT	9				
Q918F1					
ID	Q918F1	PRELIMINARY;	PRT;	1246	AA.
AC	Q918F1;				
DT	01-OCT-2000	(TReMBLrel. 15, Created)			

[illegible]

RESULT		12		
C9JJ07	PRELIMINARY;		PTRT;	579 AA.
ID	G9JJ07			
AC	Q9JJ07;			
ET.	01-OCT-2000 (TEMBLrel. 15, Created)			
DY	01-OCT-2000 (TEMBLrel. 15, Last sequence update)			
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation change)			
DE	RECEPTOR-TYPE PROTEIN TYROSINE PHOSPHATASE [FRAGMENT].			
CS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
NCBJ_TaxID=10090;	[1]			
RN	SEQUENCE FROM N.A.			
RP				
RC	TISSUE-VASCULAR ENDOTHELIUM;			
RA	fachinger G., Deutsch U., Risau W.;			
RT	"Functional interaction of vascular endothelial protein tyrosine			
RL	phosphatase with the angiotensin receptor tie-2";			
RF	Oncogene 0:0-0(1999).			
DR	EMBL; AF157628; AAF80346.1; -.			
IPRPro:	IPRO00387; Tyr phosphatase.			

	CARBOHYD	82	82	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	93	93	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	104	104	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	142	142	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	172	172	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	192	192	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	231	231	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	258	258	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	278	278	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	342	342	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	351	351	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	376	376	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	391	391	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	396	396	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	413	413	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	431	431	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	501	501	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	525	525	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	536	536	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	582	582	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	603	603	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	618	618	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	628	628	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	637	637	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	666	666	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	669	669	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	761	761	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	772	772	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	784	784	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	790	790	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	824	824	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	910	910	N-LINKED	(GLCNAC .)	(POTENTIAL) .
FT	CARBOHYD	937	937	N-LINKED	(GLCNAC .)	(POTENTIAL) .
SQ	SEQUENCE	I337 AA; 145801 MW; BA7EA3BE3296ICEA CRC64;				

Query Match 35.5%; Score 239; DB 4; Length 1337;
Best Local Similarity 38.1%; Pred. No. 3.1e-18;
Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;

```

QY      1 DFWGMMWENNVCVIAITREIEGGVIKCCSYWPVSUKEPLEKFHFHVLLENFQITQYFYI 60
        ||| :|| : | |:| :| | :| | | | : : | : : | : : | : | : | : | : |
DB    1122 DFRVMWXNRNVAIMLTCTCVEOGRTKCEBYP--SKQAQDYGDITVAMTSEIVLPEWTI II79
               ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY      61 RIFOIKVKSTGSHKVHLQFIKWDPHTGPASVDFFIKY---VR-YVRKSHITGPLLVHC 116
        ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB    1180 RDFTVNQIOTSHPRIQRPHFTSWPDHGVPDPTDLLINFRYLVRDKMKSPESPILVHC 1239
                ||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY      i.i7 TAGVGR 122
        :|||||
DB     1240 SAGVGR 1245
              ||| |

RESULT { 14
QQ0947 ID Q90947 PRELIMINARY; PRT; 832 AA.
AC AC Q90947;
DT DT 01-NOV-1996 (TREMBLrel_01, Created)
DE DE 01-NOV-1996 (TREMBLrel_01, Last sequence update)
LE LE 01-JUN-2001 (TREMBLrel_17, Last annotation update)
OS OS PHOSPHORYLSYL PHOSPHATASE (EC 3.1.3.48).
   OC Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN RN NCBI_TaxID=9031;
RC RC [1]
RX RX SEQUENCE FROM N.A.
RA RA TISSUE=EMBRYO;
RT RT MEDLINE=941122757; PubMed=8293038;
RA Rowley R., Lee J.M., Corbell H.B., Charbonneau R., Jue K.,
RT Bankort D.L., Branton P.E.;
RT "Isolation of chicken phosphotyrosyl phosphatase cDNA sequences and

```

RT identification of a brain-specific species related to human PTP

```
RT zeta.;
RL Cell. Mol. Biol. Res. 39:209-219(1993).
DR EMBL; L27625; AAA49015.1; -.
DR HSP; P18052; IYFO.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
SQ SEQUENCE 832 AA; 93297 MW; 4A61F365B8D794F3 CRC64;

Query Match 35.4%; Score 238; DB 13; Length 832;
Best Local Similarity 37.0%; Pred. No. 2.3e-18;
Matches 50; Conservative 24; Mismatches 43; Indels 18; Gaps 5;

Qy 1 DFWGMWNNCNVIAITREIEGGVIKCCSYWPVSLKEPFEKHFHVLLENFQITQYFVI 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 DFRMIWEHNVEVIMITNLLKGRKCDQYPAEGSE--EYGNFLVTQKSVHVLAYTV 384

Qy 61 RIFOI----VKKSTGKSHS----VKHLQFIKWPDHGTPTASVDFFIKYVRYVRKS-----H 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 RNFTLRNTKIKKSGOKGRSGRVVTOYHYTQWPDGVP---EYTLPLVTFVRKASHAKRH 441

Qy 108 ITGPLLHVCTAGVGR 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 AVGPVVVHCSAGVGR 456

RESULT 15
Q9UDR6
ID Q9UDR6 PRELIMINARY; PRT; 1648 AA.
AC Q9UDR6; 2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE WUGSC:H_DJ1049N15.1 PROTEIN (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX Drone K., Le T.;
RT "The sequence of Homo sapiens PAC clone RP5-1049N15.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX Waterston R.H.;
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX Waterston R.;
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC006020; AAF03527.1; -.
DR HSP; P18052; IYFO.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
```

```
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON TER .1
SQ SEQUENCE 1648 AA; 180082 MW; D4D10F9BDD4A6A55 CRC64;

Query Match 35.4%; Score 238; DB 4; Length 1648;
Best Local Similarity 37.0%; Pred. No. 5e-18;
Matches 50; Conservative 24; Mismatches 43; Indels 18; Gaps 5;

Qy 1 DFWGMWNNCNVIAITREIEGGVIKCCSYWPVSLKEPFEKHFHVLLENFQITQYFVI 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1343 DFRMIWEHNVEVIMITNLLKGRKCDQYPAEGSE--EYGNFLVTQKSVQVLAITYTV 1200

Qy 61 RIFOI----VKKSTGKSHS----VKHLQFIKWPDHGTPTASVDFFIKYVRYVRKS-----H 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1201 RNFTLRNTKIKKSGOKGRSGRVVTOYHYTQWPDGVP---EYSLPLVTFVRKAAAYAKRH 1257

Qy 108 ITGPLLHVCTAGVGR 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 AVGPVVVHCSAGVGR 1272

Search completed: March 30, 2002, 08:24:43
Job time: 655 sec
```


THIS PAGE BLANK (USPTO)